



# Learning Structure from Visual Semantic Features and Radiology Ontology for Lymph Node Classification on MRI

Yingying Zhu<sup>3</sup>(✉), Shuai Wang<sup>1</sup>, Qingyu Chen<sup>2</sup>, Sungwon Lee<sup>1</sup>, Thomas Shen<sup>1</sup>, Daniel C. Elton<sup>1</sup>, Zhiyong Lu<sup>2</sup>, and Ronald M. Summers<sup>1</sup>(✉)

<sup>1</sup> Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Clinical Center, National Institutes of Health, Bethesda, MD 20892, USA  
rms@nih.gov

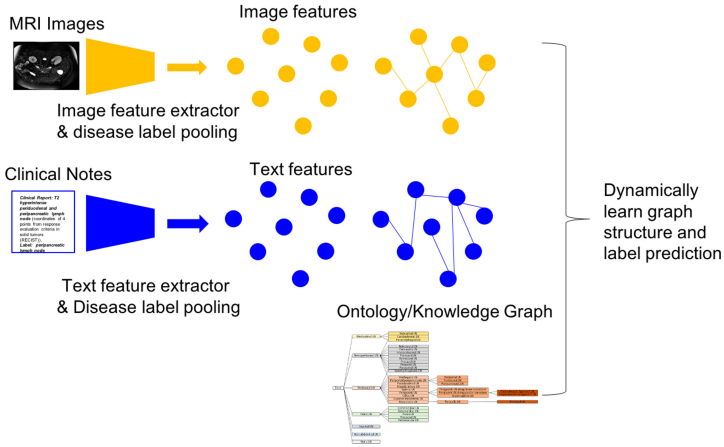
<sup>2</sup> National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA

<sup>3</sup> Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019, USA  
yingying.zhu@uta.edu

**Abstract.** Medical image classification (for example, lesions on MRI scans) is a very challenging task due to the complicated relationships between different lesion sub-types and expensive cost to collect high quality labelled training datasets. Graph model has been used to model the complicated relationship for medical imaging classification successfully in many previous work. However, most existing graph based models assumed the structure is known or pre-defined, and the classification performance severely depends on the pre-defined structure. To address all the problems of current graph learning models, we proposed to jointly learn the graph structure and use it for classification task in one framework. Besides imaging features, we also use the disease semantic features (learned from clinical reports), and predefined lymph node ontology graph to construct the graph structure. We evaluated our model on a T2 MRI image dataset with 821 samples and 14 types of lymph nodes. Although this dataset is very unbalanced on different types of lymph nodes, our model shows promising classification results on this challenging datasets compared to several state of art methods.

## 1 Introduction

Accurate classification of lymph nodes is of very important clinical meaning for the diagnosis or prognosis of numerous diseases such as metabolises cancer and can be used as assistance for the radiologist to locate abnormality and diagnosis different diseases. Developing machine learning methods to automatically identify different lymph nodes from MRI scans is a very challenging task due to the similar morphological structures of lymph nodes, very complicated relationship between different types of lymph nodes and the expensive cost to collect large-scale labelled image datasets. Labelling lymph node types from the MRI scans is



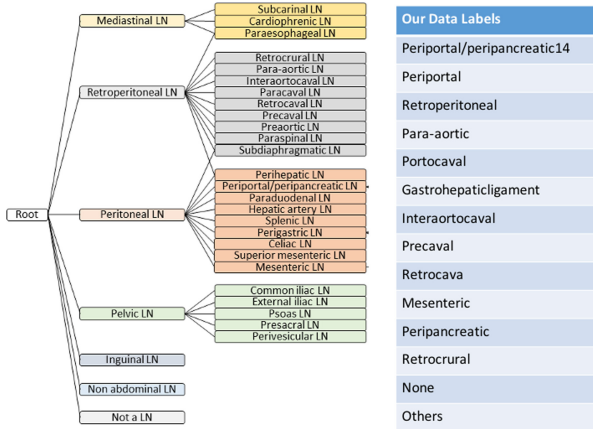
**Fig. 1.** The framework of the classical graph based classification.

very time consuming and expensive since it requires the well trained radiologist to look into each MRI slices, therefore, it is not realistic to collect a large size labeled dataset to train a deep learning system.

In this work, instead of manually labelling images, we extracted lymph node key words from clinical notes associated with MRI scans by experienced radiologists and used the extracted key words as classification labels for the corresponding MRI images. We extracted 14 different types of lymph nodes from the clinical reports (as shown in Fig. 3) from 821 T2 MRI key slices. It is worth noting that our dataset is highly unbalanced on different types of lymph nodes (some lymph node has more than 80 training images and some lymph node only have less than 10 training images). Considering the unbalanced small size training dataset, it is very challenging to train the classification model with high accuracy and generalizability.

Motivated by recent works on successfully using semantic information for image classification and image captioning problems [10, 11], we proposed to leverage the semantic features learned from clinical notes along with a predefined lymph node ontology graph (as shown in Fig. 2) by radiologist for the lymph node classification. We proposed to learn a semantic feature embedding on the clinical reports and used it to learn the semantic relationship between different lymph node types. Based on this semantic embedding space, we also combined it with the ontology graph (shown in Fig. 1 (b)) to construct a knowledge graph to guide our classification task. Besides, most existing graph models assume that the graph structure is known or predefined. When graph structure is unknown, they usually use K-nearest neighbor to construct a graph structure from image features and use it for downstream image classification task [2, 8]. Defining graph structure is critical for the downstream graph node classification task, recent work shows that learning a graph structure can significantly boost the performance of graph based classification tasks [12]. In this work, we also proposed a

joint graph structure learning and classification framework with prior information from semantic features (learned from clinical reports) and radiologist defined lymph node ontology graph. We evaluated our proposed model on lymph node classification in T2 MRI scans and show consistent improvement compared to several state-of-the-art methods given a small and imbalanced training dataset.



**Fig. 2.** The predefined ontology graph of a lymph node structure by an experienced radiologist. We use this ontology graph as a prior to construct our knowledge graph. The 14 labels on our dataset is also shown here.

## 2 Method

**Graph Learning for Classification.** We define a graph  $G = (\mathbf{V}, E)$ , where  $\mathbf{V}$  is the set of vertices or nodes (we will use nodes throughout the paper), and  $E$  is the set of edges. Let  $\mathbf{v}_i \in \mathbf{V}$  denote a node and  $e_{ij} = (\mathbf{v}_i, \mathbf{v}_j) \in E$  to denote an edge pointing from  $\mathbf{v}_j$  to  $\mathbf{v}_i$ . The adjacency matrix  $\mathbf{\Lambda}$  is a  $n \times n$  matrix with  $\Lambda_{ij} = 1$  if  $e_{ij} \in E$  and  $\Lambda_{ij} = 0$  if  $e_{ij} \notin E$ . It is worth noting that we only consider about the un-directed graph in this work, thus  $\mathbf{\Lambda}$  is symmetric. It is straightforward to extend our frame work to the directed graph with different graph structure regularization terms. The node of graph has node feature  $x_i \in \mathcal{R}^{d \times 1}$ ,  $X = [x_1, \dots, x_n] \in \mathbb{R}^{n \times d}$  is a node feature matrix for all nodes in the graph. Let  $\mathbf{y}_i$  denotes the class labels of node  $x_i$  and  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$  denotes the node class labels of all different nodes on the graph. Conventional graph learning methods usually define a function  $f(\mathbf{X}, \mathbf{\Lambda}) = \mathbf{Y}$  with input as node feature vector matrix  $\mathbf{X}$  and graph adjacency matrix  $\mathbf{\Lambda}$  to estimate the class labels  $\mathbf{Y}$  for each node.  $f$  can be a simple linear model or a graph neural network model. The classic graph learning solves the following function to learn  $f$ ,

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Y}) &= \|\mathbf{Y} - f(\mathbf{X}, \mathbf{\Lambda})\|_p + \lambda f(\mathbf{X}, \mathbf{\Lambda})^T \mathbf{f}(\mathbf{X}, \mathbf{\Lambda}) \\ \text{s. t. } & \mathbf{t} = \mathbf{I} - \mathbf{\Lambda}, \end{aligned} \quad (1)$$

where  $\|\cdot\|_p$  represents  $L_p$  norm, the first term is the label prediction term and the second term is the laplacian constraint term,  $\mathbf{I}$  is the identity matrix, is the laplacian matrix of graph. If using a multi-layer deep graph neural network model for  $f$

$$\begin{aligned} \mathbf{L}(\mathbf{X}, \mathbf{Y}) &= \|\mathbf{Y} - f(\mathbf{X}, \mathbf{\Lambda})\|_p + \lambda f(\mathbf{X}, \mathbf{\Lambda})^T \mathbf{f}(\mathbf{X}, \mathbf{\Lambda}) \\ \text{s. t. } &= \mathbf{I} - \mathbf{\Lambda} \end{aligned} \quad (2)$$

These graph learning methods always assumed that the graph structure is predefined (the adjacency matrix  $\mathbf{\Lambda}$  is given), which is not applicable for many cases. For example, the connection or similarity between different patients or different types of lymph nodes is hard to defined. In practice, many previous works just use a K-nearest neighbor to extract the graph adjacency matrix and use it for label propagation on the graph. Recent works have shown that defining an optimal graph structure is crucial for the label classification task on the graph and jointly optimizing label propagation and learning graph structures can significantly improve the performance and generalizability of graph models [2].

**Jointly Learning Graph Structure and Classification.** We propose to learn the graph structure  $\mathbf{\Lambda}$  jointly with graph label propagation, let  $\mathbf{\Lambda} = g(\mathbf{X})$  be the function to learn the graph structure, we define the new loss function as,

$$\begin{aligned} \mathbf{L}(\mathbf{X}, \mathbf{\Lambda}, \mathbf{Y}) &= \|\mathbf{Y} - f(\mathbf{X}, \mathbf{\Lambda})\|_p + \lambda_0 f(\mathbf{X}, \mathbf{\Lambda})^T \mathbf{f}(\mathbf{X}, \mathbf{\Lambda}) \\ &+ \lambda_g \|\mathbf{\Lambda} - g(\mathbf{X})\|_p, \text{ s. t. } = \mathbf{I} - \mathbf{\Lambda} \end{aligned} \quad (3)$$

There are many ways to construct the graph structure learning  $g$  function, for example, previous work [2] has tried to use Bernoulli sampling to learning the graph structure from discrete nodes, or using sparse and low rank subspace learning to learn the graph structure. We follow the [1, 6, 13] to use the sparse and low rank subspace learning in our work. However, it is straight forward to extend our work with different graph structure learning methods. We reformulated our objective function as,

$$\begin{aligned} \mathbf{L}(\mathbf{X}, \mathbf{\Lambda}, \mathbf{Y}) &= \|\mathbf{Y} - f(\mathbf{X}, \mathbf{\Lambda})\|_p + \lambda_0 f(\mathbf{X}, \mathbf{\Lambda})^T \mathbf{f}(\mathbf{X}, \mathbf{\Lambda}) \\ &+ \lambda_1 \|\mathbf{\Lambda}\|_1 + \lambda_2 \|\mathbf{\Lambda}\|_*, \text{ s. t. } = \mathbf{I} - \mathbf{\Lambda}, \mathbf{X} = \mathbf{X}\mathbf{\Lambda}, \end{aligned} \quad (4)$$

where  $\mathbf{\Lambda}$  is constrained to be low rank using the nuclear norm  $\|\cdot\|_*$  and sparse using  $L_1$  norm.

**Predefined Knowledge Graph.** Many previous works have shown that expert knowledge is tremendously helpful for medical data analysis especially when labelled training data-size is small. In this work, we only have 821 labelled MRI slices as training data, which is very small for training a graph neural network. In order to further improve our model, we extracted a knowledge graph from radiologist labelled lymph node ontology graph as shown in Fig. 2. For all labelled training images/nodes, we will construct in-directed edges between them, if they are connected in the predefined ontology graph.

Besides the ontology graph, since we can access the clinical report associated with each MRI slice in our dataset, we also use the MRI reports (shown in Fig. 3) to train a report classification model using pre-trained BERT [4] and apply it to extract a semantic vector by attention pooling for different lymph nodes. Based on the similarity between these semantic features and the ontology graph, we constructed a knowledge graph for node label classification. Denote the adjacency matrix of the knowledge graph as  $A_{kg}$ , our problem is further formulated as,

$$\begin{aligned} L(\mathbf{X}, \mathbf{\Lambda}, \mathbf{Y}) = & \|\mathbf{Y} - f(\mathbf{X}, \mathbf{\Lambda})\|_p + \lambda_0 f(\mathbf{X}, \mathbf{\Lambda})^T \mathbf{f}(\mathbf{X}, \mathbf{\Lambda}) \\ & + \lambda_1 \|\mathbf{\Lambda}\|_1 + \lambda_2 \|\mathbf{\Lambda}\|_*, \text{ s. t. } = \mathbf{I} - (\mathbf{\Lambda} + \beta \mathbf{\Lambda}_{kg}), \mathbf{X} = \mathbf{X}\mathbf{\Lambda}, \end{aligned} \quad (5)$$

where  $\beta$  is a hyper-parameter to learn for adding the knowledge graph as prior in this problem. In order to solve Eq. 6, we use a Lagrange multiplier to add the constraints to the objective function,

$$\begin{aligned} L(\mathbf{X}, \mathbf{\Lambda}, \mathbf{Y}) = & \|\mathbf{Y} - f(\mathbf{X})\|_p + \lambda_0 f(\mathbf{X})^T (\mathbf{I} - (\mathbf{\Lambda} + \beta \mathbf{\Lambda}_{kg})) \\ & \mathbf{f}(\mathbf{X}) + \lambda_1 \|\mathbf{\Lambda}\|_1 + \lambda_2 \|\mathbf{\Lambda}\|_* + \lambda_3 \|\mathbf{X} - \mathbf{X}\mathbf{\Lambda}\|_2^2 \end{aligned} \quad (6)$$

**Graph Convolutional Neural Networks.** Graph convolutional neural networks have shown great successes on many applications [3, 14], our proposed framework can be easily combined with graph convolutional neural networks. Let  $\mathbf{H}$  denotes the hidden states of a graph neural network,  $\mathbf{W}_l$  as the weights for the hidden layer  $l$ , a standard graph convolutional neural network can be formulated  $\mathbf{H}_{l+1} = f(\mathbf{W}_l, \mathbf{\Lambda}, \mathbf{H}_l)$ , where the input is the image feature matrix  $\mathbf{X}$  and output is the predicted labels  $\mathbf{Y}$ , thus,  $\mathbf{H}_0 = \mathbf{X}$ ,  $\mathbf{H}_{last} = \mathbf{Y}$ ,  $\mathbf{H} = [\mathbf{H}_0, \dots, \mathbf{H}_{last}]$ . We can rewrite Eq. 7 using graph convolutional neural networks as,

$$\begin{aligned} L(\mathbf{X}, \mathbf{\Lambda}, \mathbf{Y}, \mathbf{W}, \mathbf{H}) = & \|\mathbf{Y} - f(\mathbf{X}, \mathbf{W}, \mathbf{\Lambda}, \mathbf{H})\|_p \\ & + \lambda_0 f(\mathbf{X}, \mathbf{W}, \mathbf{\Lambda}, \mathbf{H})^T (\mathbf{I} - (\mathbf{\Lambda} + \beta \mathbf{\Lambda}_{kg})) \mathbf{f}(\mathbf{X}, \mathbf{W}, \mathbf{\Lambda}, \mathbf{H}) \\ & + \lambda_1 \|\mathbf{\Lambda}\|_1 + \lambda_2 \|\mathbf{\Lambda}\|_* + \lambda_3 \|\mathbf{X} - \mathbf{X}\mathbf{\Lambda}\|_2^2. \end{aligned} \quad (7)$$

**Optimization Method.** It is trivial to optimize both the graph structure and node classification on the same dataset. In order to solve Eq. 7, we proposed a bi-level optimization method to learn the graph structure  $\mathbf{\Lambda}$ , GNN model weight  $\mathbf{W}$  and labels  $\mathbf{Y}$  jointly. The graph structure adjacency matrix  $\mathbf{\Lambda}$  is learned on the validation dataset, and the GNN model weight  $\mathbf{W}$  is learned on the training dataset. Hyperparameters  $\lambda_0, \lambda_1, \lambda_2, \lambda_3, \beta$  are also learned on the validation dataset too. The detailed optimization algorithm is shown in Algorithm 1.

---

**Algorithm 1:** Optimization algorithm for jointly learn  $\Lambda, \mathbf{W}$ 

---

**Result:** Graph Adjacency Matrix  $\Lambda$ , Graph Neural Network weights  $\mathbf{W}$ 

```

1 Initialize parameter  $\mathbf{W}$  randomized,  $\mathbf{H}_0 = \mathbf{X}, \mathbf{H}_{last} = \mathbf{Y}, \lambda_0, \lambda_1, \lambda_2, \lambda_3$ ;
2  $i = 1$ , max iteration of outer loop  $max_{iter} = 1e5$ ;
3 max iteration of inner loop  $max_k = 1e3$ ;
4 learning rate  $\rho = 1e - 3$ ;
5 while  $i < max_{iter}$  do
6    $\mathbf{W}^{i+1} = \mathbf{W}^i + \mu_w \frac{\partial \mathbf{L}^i(\mathbf{X}, \Lambda^k, \mathbf{W}^i, \mathbf{Y}^i, \mathbf{H}^i)}{\mathbf{W}}$ ;
7    $\mu_w = \frac{\partial^2 \mathbf{L}^i(\mathbf{X}, \Lambda^k, \mathbf{W}^i, \mathbf{Y}^i, \mathbf{H}^i)}{\mathbf{W}^2}$ ;
8    $i = i + 1$ ;
9    $k = 1$ ;
10  while  $k < max_k$  (Validation Dataset) do
11     $\Lambda^{k+1, i} = \Lambda^{k, i} + \mu_A \frac{\partial \mathbf{L}^{k, i}(\mathbf{X}, \Lambda^{k, i}, \mathbf{W}^i, \mathbf{Y}^i, \mathbf{H}^i)}{\Lambda}$ ;
12     $\mu_A = \frac{\partial^2 \mathbf{L}^i(\mathbf{X}, \Lambda^k, \mathbf{W}^i, \mathbf{Y}^i, \mathbf{H}^i)}{\Lambda^2}$ ;
13     $\mu_A = \rho \mu_A + \mu_A$ ;
14     $k = k + 1$ ;
15  end
16   $\mu_W = \rho \mu_W + \mu_W$ ;
17 end

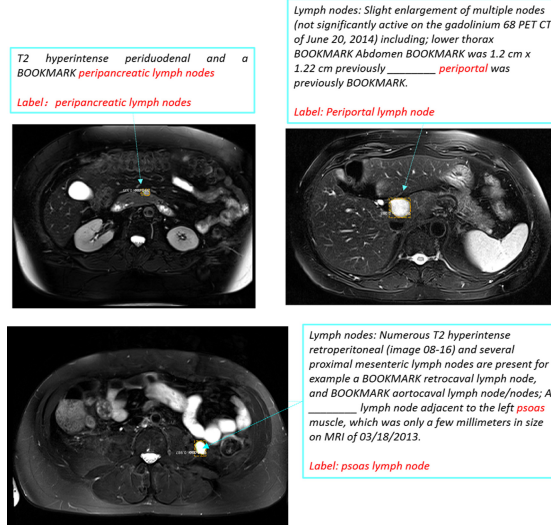
```

---

### 3 Experiments

**Dataset.** For model development and validation, we collected large-scale MRI studies from \*\*, performed between Jan 2015 to Sept 2019 along with their associated radiology reports. The majority (63%) of the MRI studies were from the oncology department. This dataset consists of a total of 821 T2-weighted MRI axial slices from 584 unique patients. The lymph node labels were extracted by a radiologist with 8 years of post-graduate experience. The study was a retrospective study and was approved by the Institutional Review Board with a waiver of informed consent. This dataset comprised the reference (gold) standard for our evaluation and comparative analysis.

**Benchmark Methods.** We implemented several benchmark methods in our experiments. 1) Support Vector Machine (SVM) [7]: applying classical SVM on the extracted multi-scale bounding box features. 2) Structured SVM: constraining the support vector machine to output structural labels constrained by the knowledge graph structure in Fig. 2. 3) Standard Simple Graph Model (SG); 4) SG with Graph Structure Learning (SG+SL); 5) SG with SL and Predefined Knowledge Graph (SG+SL+KG); 6) Deep Neural Graph (GCN); 7) GCN with Graph Structure Learning (GCN+SL); 8) GCN with SL and Predefined Knowledge Graph (GCN+SL+KG); 9) Deep Neural Hyper-Graph (HGCN)[9]; 10) HGCN with Graph Structure Learning (HGCN+SL); 11) HGCN with SL and Predefined Knowledge Graph (HGCN+SL+KG). We use the same lymph



**Fig. 3.** An example of MRI slice with the labelled bounding boxes for lymph-nodes and graph-cut based detailed annotations of lymph nodes, linked clinical report sentences and radiologist labelled lymph node names.

node image feature embedding framework for all competing methods since we want to show the different classification performance between our methods and all other benchmark methods.

**Experiment Setting and Data Processing.** We divided the dataset into 10 folds, use one fold as validation dataset and two folds as testing dataset, the left seven folds are used as the training dataset. We run the cross validation for 10 times and report the averaged top-k ( $k = 1, 2, 3$ ) classification of accuracy for different types of lymph nodes, F1-score and AUC of binary classification performance. In our dataset, we have the access to the clinical report of each MRI scan. The radiologist describes the lymph node information including the labels, size measurements, and slice numbers in a sentence with hyperlink (called bookmark) referred to the related MRI slices. The radiologist defines a bookmark as a hyperlink connection between the annotation in the image and then writes description in the report. We have one experienced radiologist to extract the lymph node labels from the bookmark linked sentences and use them as the ground truth labels for the lymph nodes in the connected MRI slices.

The size of lymph nodes are measured by four points at maximum dimension of lymph nodes or two points at the maximum dimension of lymph nodes. Based on these key points, we extracted multi-scale bounding boxes around the lymph nodes and extract the features in these bounding boxes using pretrained CNN model on MRI slices. We further use graph-cut to extract the fine contours of the lymph nodes and extract the cut lymph nodes using pretrained CNN model. We concatenated all these multi-scale bounding box features and lymph

**Table 1.** The top-k ACC & top-k F1 score of multi-class lymph node classification performance by different competing methods

Method	top-1 ACC	top-2 ACC	top-3 ACC	top-1 F1	top-2 F1	top-3 F1
SVM	0.72	0.78	0.85	0.70	0.76	0.83
Structured SVM	0.75	0.81	0.87	0.74	0.80	0.85
SG	0.73	0.80	0.86	0.72	0.79	0.84
SG+SL	0.76	0.82	0.88	0.75	0.81	0.87
SF+SL+KG	0.79	0.85	0.90	0.78	0.84	0.88
GCN	0.75	0.81	0.87	0.74	0.80	0.86
GCN+SL	0.81	0.84	0.89	0.80	0.83	0.88
GCN+SL+KG	<b>0.83</b>	<b>0.86</b>	<b>0.91</b>	<b>0.82</b>	<b>0.86</b>	<b>0.90</b>
Hyper-GCN	0.76	0.83	0.88	0.75	0.82	0.87
Hyper-GCN+SL	0.83	0.87	0.91	0.82	0.86	0.90
Hyper-GCN+SL+KG	<b>0.85</b>	<b>0.89</b>	<b>0.93</b>	<b>0.84</b>	<b>0.88</b>	<b>0.92</b>

node features and use it as the graph node feature representation. The length of the concatenated multi-scale feature vector is 25088. We use the pre-trained bioBERT to train the clinical notes classification and label attention to extract the semantic label embedding. We used about more than 28000 sentences of de-identified clinical reports from \*\* hospital and use it to embed the distance between different lymph node names in our dataset. Based on the semantic distance between different lymph nodes, we constructed a semantic embedding graph and combine it with the predefined ontology graph shown in Fig. 2 to refine the final label prediction results.

**Quantitative Results.** We compared our proposed model to several benchmarks and Table 1 shows the top-k mean accuracy and F1 score of 14 classification results on the 10-fold cross validation. Top-k accuracy has been broadly used for multi-class classification performance measurement in previous work [5]. We show that the simple graph model generally outperforms SVM methods, and the structured SVM improves classical SVM about  $> 0.03$  on both accuracy and F1 score by adding the structured constraints on different classes (extracted from the pre-defined ontology graph). Learning graph structure improves the top-k accuracy of the simple graph model by  $> 0.03$  and F1 score about  $> 0.02$ . Using knowledge graph improves the simple graph mode further about  $> 0.03$  on both top-k accuracy and F1 score. The convolutional neural graph also improves the classification top-k accuracy and F1 score consistently compared to the simple graph model. Learning graph structure and using knowledge graph under convolutional neural graph framework, our proposed model achieves the best performance and shows about 0.91 on top-3 accuracy and 0.90 on top-3 F1 score. We also combine the graph learning method with convolutional hyper-graph model and show that it improves the accuracy and F1-score  $> 2\%$  compared to convolutional graph model. The best top-3 accuracy and F1 score achieved by hyper-graph model is 93% and 0.92%.



**Acknowledgment.** This research was supported in part by the Intramural Research Program of the National Institutes of Health, Clinical Center and National Library of Medicine.

## References

1. Chen, J., Yang, J.: Robust subspace segmentation via low-rank representation. *IEEE Trans. Cybern.* **44**(8), 1432–1445 (2014)
2. Franceschi, L., Niepert, M., Pontil, M., He, X.: Learning discrete structures for graph neural networks. In: *Proceedings of the 36th International Conference on Machine Learning* (2019)
3. Hamilton, W.L., Ying, R., Leskovec, J.: Representation learning on graphs: Methods and applications. *CoRR abs/1709.05584* (2017)
4. Lee, J., et al.: BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **36**(4), 1234–1240 (2019)
5. Lu, J., Xu, C., Zhang, W., Duan, L.Y., Mei, T.: Sampling wisely: Deep image embedding by top-k precision optimization. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019
6. Sui, Y., Wang, G., Zhang, L.: Sparse subspace clustering via low-rank structure propagation. *Pattern Recogn.* **95**, 261–271 (2019)
7. Tsochantaridis, I., Joachims, T., Hofmann, T., Altun, Y.: Large margin methods for structured and interdependent output variables. *J. Mach. Learn. Res.* **6**, 1453–1484 (2005)
8. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. *CoRR abs/1901.00596* (2019)
9. Yadati, N., Nimishakavi, M., Yadav, P., Louis, A., Talukdar, P.P.: Hypergn: Hypergraph convolutional networks for semi-supervised classification. *CoRR abs/1809.02589* (2018). <http://arxiv.org/abs/1809.02589>
10. Zareian, A., Karaman, S., Chang, S.: Bridging knowledge graphs to generate scene graphs. In: *CVPR* (2020)
11. Zhang, D., et al.: Knowledge graph-based image classification refinement. *IEEE Access* **7**, 57678–57690 (2019)
12. Zhou, Y., Sun, Y., Honavar, V.G.: Improving image captioning by leveraging knowledge graphs. *CoRR abs/1901.08942* (2019)
13. Zhu, X., Zhang, S., Li, Y., Zhang, J., Yang, L., Fang, Y.: Low-rank sparse subspace for spectral clustering. *IEEE Trans. Knowl. Data Eng.* **31**(8), 1532–1543 (2019)
14. Zhu, Y., Zhu, X., Kim, M., Yan, J., Kaufer, D., Wu, G.: Dynamic hyper-graph inference framework for computer-assisted diagnosis of neurodegenerative diseases. *IEEE Trans. Med. Imaging* **38**(2), 608–616 (2019)