



# Edge-Variational Graph Convolutional Networks for Uncertainty-Aware Disease Prediction

Yongxiang Huang<sup>(✉)</sup> and Albert C. S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China  
{yhuangch, achung}@cse.ust.hk

**Abstract.** There is a rising need for computational models that can complementarily leverage data of different modalities while investigating associations between subjects for population-based disease analysis. Despite the success of convolutional neural networks in representation learning for imaging data, it is still a very challenging task. In this paper, we propose a generalizable framework that can automatically integrate imaging data with non-imaging data in populations for uncertainty-aware disease prediction. At its core is a learnable adaptive population graph with variational edges, which we mathematically prove that it is optimizable in conjunction with graph convolutional neural networks. To estimate the predictive uncertainty related to the graph topology, we propose the novel concept of Monte-Carlo edge dropout. Experimental results on four databases show that our method can consistently and significantly improve the diagnostic accuracy for Autism spectrum disorder, Alzheimer’s disease, and ocular diseases, indicating its generalizability in leveraging multimodal data for computer-aided diagnosis.

**Keywords:** Population-based disease prediction · Graph neural network · Deep learning

## 1 Introduction

Integrating imaging data with non-imaging data for disease diagnosis is an essential task in clinics. In recent years, the increasing volume of digitalized multimodal data has raised the need for computational models with the capability of exploiting different modalities automatically for improving prediction accuracy and discovering new biomarkers to study the disease mechanism (e.g., Alzheimer’s disease) [23]. Despite the success of convolutional neural networks (CNNs) in medical images [8, 14], exploiting both imaging and non-imaging data in populations in a unified model can be challenging. Multimodal learning-based approaches usually summarize features of all modalities with a deep neural network for disease classification [26], which ignore the interaction and association

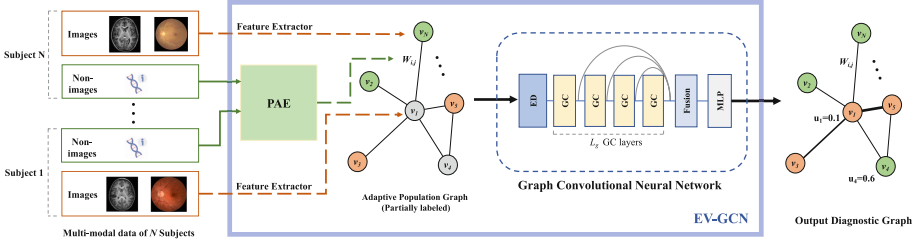
between subjects in a population. Graphs provide a natural way to represent the population data and enable the use of powerful tools such as clustering algorithms for disease analysis. Moreover, recent studies on graph convolutional neural networks (GCNs) [11, 24] have extended the theory of signal processing on graphs [20] to complement the representation learning limitation of CNNs on irregular graph data.

In this work, we present a generalizable framework to automatically integrate multimodal data in populations for disease prediction with uncertainty estimation. Our contributions include: i) proposing a novel adaptive population graph model for representing multimodal features and associations for subjects and mathematically showing that it can be optimized in conjunction with spectral GCNs, which makes semi-supervised learning with GCNs generalizable for medical databases, ii) proposing Monte-Carlo edge dropout for estimating the predictive uncertainty with respect to the graph topology, which is new and extendable for graph neural networks, iii) designing a well-regularized spectral graph convolutional network for population-based disease prediction, alleviating the over-smoothing problem, and iv) extensively evaluating our method and recent multimodal learning methods on four challenging databases, which shows the proposed method can significantly improve the diagnostic performance for brain analysis and ocular diseases. (To the best of our knowledge, it is also the first study of GCNs on population-based ocular disease analysis.).

**Related Work.** Recent studies [11, 13, 24] have shown that a graph can serve as a regularizer on node classification tasks for semi-supervised learning with GCNs. However, these methods are evaluated on graph benchmarks where the associations between nodes are inherently defined in the data (e.g., a citation dataset [11]). Contrastively, in the medical domain, the associations between subjects (i.e., nodes) are usually uncertain and multifaceted, especially for multimodal databases. The absence of well-defined relation between two nodes leads to the uncertainty of graph topology, making it hard to adopt GCNs for population-based disease diagnosis. A few recent studies investigated the construction of affinity graphs by computing the correlation distance between subjects for brain analysis [9, 17]. These methods require to manually tune distance thresholds for different modalities for the graph construction, which can heavily fluctuate the performance, leading to a lack of generalizability.

## 2 Methods

In this section, we present the proposed model, namely Edge-Variational GCN (EV-GCN), for incorporating multimodal data for disease prediction. The overview of the pipeline is depicted in Fig. 1. EV-GCN accepts the imaging features and non-imaging data of  $N$  subjects and constructs an adaptive population graph with partially labeled nodes and variational edges, followed by the proposed spectral graph convolutional network with edge dropout for learning to estimate the diagnostic value and the uncertainty for each testing subject.



**Fig. 1.** Overview of the proposed method. PAE: pairwise association encoder. ED: edge dropout. GC: graph convolution. Fusion: vertex-wise concatenation. Colors in the graphs: green and orange - labeled diagnostic values (e.g., healthy or diseased), grey: unlabeled.  $u_i$ : predictive uncertainty for subject  $i$ . (Color figure online)

## 2.1 Edge-Variational Population Graph Modeling

Given the observation of  $N$  subjects composed of imaging and non-imaging data, let us consider constructing a population graph  $G = (V, E, W)$ , where  $V$  is a finite set of vertices with  $|V| = N$ ,  $E \subseteq V \times V$  is a set of edges with  $W$  being the edge weights. To associate a vertex  $v \in V$  with the diagnostic features of a subject, we define the node feature  $Z_i \in \mathbb{R}^C$  as a  $C$ -dimensional feature vector extracted from the imaging data of subject  $i$ , under the observation that imaging data (e.g., histology images, functional MRI) usually provide the most important evidence for diagnosis. The modeling for the graph connectivity (i.e., edge weight) is critical for the task performance as it encodes the associations between subjects. Unlike previous methods modeling the edge weight statistically [9, 17], we propose to define the edge weight  $w_{i,j} \in W$  between the  $i$ -th and  $j$ -th vertices as a learnable function of their non-imaging measurements  $(\mathbf{x}_i, \mathbf{x}_j)$ , considering non-imaging data can provide additional information (e.g., gender, age, gene) complementary to imaging features and explain potential association. The learnable function  $f : (\mathbf{x}_i, \mathbf{x}_j) \mapsto \mathbb{R}$  is modeled by the proposed pairwise association encoder (PAE) with trainable parameters  $\Omega$  such that  $w_{i,j} = f(\mathbf{x}_i, \mathbf{x}_j; \Omega)$ .

**Pairwise Association Encoder.** The PAE starts by normalizing the multi-modal inputs  $\mathbf{x}_i$  and  $\mathbf{x}_j$  to zero-mean and unit-variance element-wisely to avoid the gradient vanishing problem, which is important in our setting as data from different modalities have various statistical properties. After normalization, we use a projection network to map each normalized input to a common latent space  $\mathbf{h}_i \in \mathbb{R}^{D_h}$  where cosine similarity can be better applied ( $D_h = 128$ ). The projection network is a multi-layer perceptron (MLP) with  $L_p$  hidden layers. We set  $L_p = 1$  in experiments and the latent feature for  $\mathbf{x}_i$  is formulated as  $\mathbf{h}_i = \Omega^{(2)}\sigma(\Omega^{(1)}\tilde{\mathbf{x}}_i + b)$ , where  $\sigma$  is a ReLU function and  $\tilde{\mathbf{x}}_i$  is the normalized input. Formally, the PAE scores the association between vertices  $i$  and  $j$  as

$$w_{i,j} = \frac{\mathbf{h}_i^\top \mathbf{h}_j}{2\|\mathbf{h}_i\|\|\mathbf{h}_j\|} + 0.5, \quad (1)$$

which gives the rescaled cosine similarity between two latent features. In training, the parameter  $\Omega$  in PAE is initialized by He Initialization. Each hidden layer is equipped with batch normalization and dropout to improve convergence and avoid overfitting. Notably, we find it beneficial and robust to define the pairwise association on the latent space rather than on the input space.

## 2.2 Spectral Graph Convolutions on Adaptive Graphs

In this section, we first present spectral graph convolutions and prove the connectivity of the proposed adaptive population graph can be optimized in conjunction with a spectral GCN, followed by presenting our GCN architecture.

A spectral convolution of a graph signal  $x$  with a filter  $g_\theta$  is defined as  $g_\theta \star x = U g_\theta U^T x$ , where  $U$  is the matrix of eigenvectors of the normalized graph Laplacian  $L = I_N - D^{-1/2} W D^{1/2}$  with  $D$  being the diagonal degree matrix (i.e.,  $L$  encodes the topological structure of graph  $G$ ). The intuition behind is that spatial graph convolutions can be computed in the Fourier domain as multiplications using graph Fourier transform (GFT) [20]. To reduce the computational cost ( $\mathcal{O}(N^2)$ ), *Chebyshev graph convolution* (ChebyGConv) [4] approximates spectral graph convolution using Chebyshev polynomials.<sup>1</sup> A  $K$ -order ChebyGConv is given by  $g_{\theta'} \star x \approx \sum_{k=0}^K T_k(\tilde{L}) \theta'_k x$ , where  $\tilde{L}$  is the rescaled graph Laplacian,  $T_k(\tilde{L}) = 2\tilde{L}T_{k-1}(\tilde{L}) - T_{k-2}(\tilde{L})$  is the Chebyshev polynomial defined recursively with  $T_0(\tilde{L}) = 1$  and  $T_1(\tilde{L}) = \tilde{L}$ , and  $\theta'_k$  are the filter parameters. As an analogy to CNNs, the polynomial term  $T_k(\tilde{L})$  acts as a  $k$ -localized aggregator, i.e., it combines the neighboring nodes that are  $k$ -step away from the central node, and  $\theta'_k$  acts as a node feature transformer.

**Claim:** *The graph connectivity modeled by PAE can be optimized in conjunction with a spectral graph convolutional model using gradient descent algorithms.*

*Proof:* Let us consider a convolution layer  $l + 1$  in a GCN and its input graph  $G^l = (V^l, E^l, W)$  with  $|V^l| = N$   $C^l$ -dimensional node feature vectors  $\mathbf{H}^l \in \mathbf{R}^{N \times C^l}$ . The convolution layer computes the output features as

$$\mathbf{H}^{l+1} = \sum_{k=0}^K T_k(\tilde{L}) \mathbf{H}^l \Theta_k^l, \quad (2)$$

where  $\Theta_k^l \in \mathbf{R}^{C^l \times C^{l+1}}$  are the filter parameters. Denote  $\mathcal{L}$  as the task-related loss function. To optimize the graph connectivity modeled by PAE by gradient descents, we need to guarantee that  $\mathcal{L}$  is differentiable w.r.t. the parameters  $\Omega$  of PAE. By chain rule, we have  $\frac{\partial \mathcal{L}}{\partial \Omega} = \frac{\partial \mathcal{L}}{\partial \mathbf{H}^l} \frac{\partial \mathbf{H}^l}{\partial W} \frac{\partial W}{\partial \Omega}$ . Both  $\frac{\partial \mathcal{L}}{\partial \mathbf{H}^l}$  and  $\frac{\partial W}{\partial \Omega}$  are derivable as they correspond to the gradients in the differentiable GCN and PAE respectively. The key is the derivative of the node feature vectors w.r.t. the

<sup>1</sup> ChebyGConv can achieve localized filtering on an irregular weighted graph with moderate computational cost, and comparatively performs well for our tasks.

input edge weights  $\frac{\partial \mathbf{H}^l}{\partial W}$ . For  $K = 1$ , based on Eq. 2, we can derive  $\frac{\partial \mathbf{H}^l}{\partial W} \Big|_{K=1} = \frac{\partial(I_N - D^{-1/2} W D^{1/2})}{\partial W}$ , which is derivable. For  $K > 1$ , since the polynomial term  $T_k(\tilde{L})$  in Eq. 2 is defined recursively for  $k > 1$ , by expanding  $T_k(\tilde{L})$  and by induction we can derive that  $\frac{\partial \mathbf{H}^l}{\partial W}$  is derivable and not a constant zero. Thus, the graph connectivity can be optimized to minimize the task loss.

**GCN Architecture.** As depicted in Fig. 1, our GCN model consists of  $L_G$  Chebyshev graph convolutional layers, each followed by ReLU activation to increase non-linearity, a fusion block, and an MLP predictor. To alleviate the over-smoothing problem [13] in deep GCNs, we propose to adopt jumping connections [25] with vertex-wise concatenation to fuse the hidden features in each depth, i.e.,  $\{\mathbf{H}^l\}_{l=1}^{L_G}$ . We find jumping connections are more effective than residual connections [7] in avoiding a performance deterioration as  $L_G$  increases. The MLP predictor consists of two  $1 \times 1$  convolutional layers (i.e., vertex-wise transformations) with 256 and  $C_k$  channels respectively, followed by a softmax function to drive a  $C_k$ -class disease probability vector for each subject. We employ cross-entropy loss on the labeled nodes to train the overall model.

### 2.3 Monte-Carlo Edge Dropout for Uncertainty Estimation

Motivated by Bayesian approximation for uncertainty estimation in CNNs [6, 10, 27], we propose *Monte-Carlo edge dropout* (MCED) to estimate the uncertainty for the constructed population graph structure.

In detail, edge dropout randomly drops a fraction of edges in the graph by placing a Bernoulli distributional mask on each  $e \in E$ , which can act as a graph data augementer to reduce overfitting [19] and increase the graph sparsity in training. For inference, similar to Monte-Carlo dropout [10] for uncertainty estimation in CNNs, MCED performs  $T$  stochastic forward passes on the GCN model under random edge dropout on the population graph, and obtain  $T$  disease probability vectors for a subject  $i$ :  $\{\mathbf{p}_i^{(t)}\}_{t=1}^T$  (We set  $T = 128$  in our experiments). We adopt the mean predictive entropy [10] as the metric to quantize the uncertainty. Formally, the uncertainty  $u_i$  for subject  $i$  is given by  $u_i = -\sum_c \mathbf{m}_{i,c} \log \mathbf{m}_{i,c}$  and  $\mathbf{m}_i = \frac{1}{T} \sum_t \mathbf{p}_i^{(t)}$ , where  $c$  corresponds to the  $c$ -th class. While MC dropout estimates the uncertainty on the neural network weights [6], MCED models the uncertainty on the graph topology, which is orthogonal.

## 3 Experiments and Results

In this section, we perform experimental evaluations of the proposed method, comparing our method with several SoTA methods for disease prediction. We consider two graph-based methods that exploit GCNs (i.e., AIG [9] and Parisot et al. [17]), a multimodal learning method (i.e., DNN-JFC [26]), and task-related unimodal methods [3, 21, 22].

### 3.1 Autism Spectrum Disorder Prediction on the ABIDE Database

**Dataset and Experimental Setting.** The Autism Brain Imaging Data Exchange (**ABIDE**) [5] publicly shares neuroimaging (functional magnetic resonance imaging (fMRI)) and the corresponding phenotypic data (e.g., age, gender, and acquisition site) of 1112 subjects, with binary labels indicating the presence of Autism Spectrum Disorder (ASD). For a fair comparison with the ABIDE state-of-the-art [3, 9], we choose the same 871 subjects consisting of 403 normal and 468 ASD individuals, use 10-fold cross-validation, and perform the same data preprocessing steps [3] to extract a  $C = 2000$  dimensional feature vector from fMRI representing a subject’s functional connectivity [3]. Phenotypic data is used to compute the pairwise association.

In our experiments, we set Cheybishev polynomial order  $K = 3$  and  $L_G = 4$ . All models are trained using Adam optimizer with learning rate 0.01, weight decay  $5 \times 10^{-5}$ , and dropout rate 0.2 for 300 epochs to avoid overfitting.

**Table 1.** Quantitative comparisons between different methods on ABIDE. MM: multi-modality,  $\times$ : only imaging data is used,  $\checkmark$ : both imaging and non-imaging data are used.  $\beta$  is a threshold for constructing a static affinity graph used in AIG.

Methods	MM	Accuracy	AUC	F1-score	#Param. (K)
Abraham et al. [3]	$\times$	66.80	-	-	-
DNN	$\times$	$73.25 \pm 3.69$	$74.16 \pm 4.64$	$74.81 \pm 4.86$	550
DNN-JFC [26]	$\checkmark$	$73.59 \pm 4.15$	$73.48 \pm 4.94$	$76.89 \pm 4.27$	635
Parisot et al. [17]	$\checkmark$	$75.66 \pm 4.69$	$81.05 \pm 6.13$	$78.85 \pm 4.66$	96
AIG [9]	$\checkmark$	$76.12 \pm 6.83$	$80.11 \pm 6.49$	$79.27 \pm 5.27$	290
AIG, $\beta = 3$ [9]	$\checkmark$	$72.10 \pm 7.12$	$75.43 \pm 8.85$	$73.55 \pm 6.94$	290
This work:					
EV-GCN	$\checkmark$	$80.83 \pm 4.92$	<b><math>84.98 \pm 5.74</math></b>	$81.24 \pm 5.76$	133
+ MCED	$\checkmark$	<b><math>81.06 \pm 4.83</math></b>	$84.72 \pm 6.27$	<b><math>82.86 \pm 5.51</math></b>	133

**Table 2.** Ablation study for this work. This table shows how different factors affect the performance of our method. The compared GCN architectures share the same depth. Plain: sequential GC layers. Inception: InceptionGCN [9].

Factor	Graph construction			GCN architecture			Edge dropout
Method	Random	Affinity	Adaptive	Plain	Inception	Ours	w/o ED
Accuracy	65.67	76.81	<b>80.83</b>	79.79	80.02	80.83	80.58
AUC	73.97	80.49	<b>84.98</b>	83.10	83.74	84.98	84.27

**Results and Analysis.** Table 1 shows the comparative results on ABIDE, where the mean and confidence interval ( $p < 0.05$ ) are computed across ten different initialisation seeds. DNN-JFC [26] summarizes the features of all modalities by jointly fully connected layers, which marginally outperforms a DNN (i.e. MLP) on the fMRI features. Comparatively, graph-based methods (Parisot, AIG, and ours) yield larger performance gains, benefiting from exploiting associations between subjects in the population graphs. The proposed method, EV-GCN+MCED, obtains an average accuracy of 81.06%, outperforming the recent SoTA method AIG [9], which employs static affinity graphs with InceptionGCN, by a margin of 4.94% accuracy with fewer parameters. We notice that the performance of AIG is highly sensitive to the threshold  $\beta$  for computing age affinity, where the best  $\beta = 2$  yields an average accuracy of 76.12%. To investigate the importance of learning an adaptive graph with variational edges, we train our GCN architecture on the same affinity graph used in AIG [9] and on a population graph with random connections. As depicted in Table 2, it results in a 4.02% (Affinity) and 15.16% (Random) accuracy drop respectively, indicating that the adaptive graph modeling is indeed key to achieving the best possible performance. Meanwhile, the effectiveness of the proposed GCN architecture and edge dropout regularization in training are ablatively validated in Table 2.

**Table 3.** Accuracy and uncertainty for models with different association (i.e. edge) inputs. We show the uncertainty as the mean value of all test subjects.

Association input	Accuracy	Uncertainty
Random noise	65.67	0.620
Gender	78.53	0.465
Gender, age	78.07	0.394
Site, gender, age	<b>81.06</b>	<b>0.307</b>

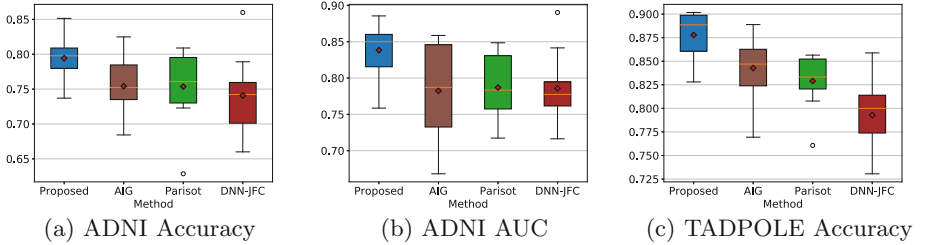
To analyze what the uncertainty estimated by MCED captures, we give the accuracy and uncertainty for models trained with different association sources (i.e. input for PAE) in Table 3. The results demonstrate that the graph uncertainty, which is approximated by averaging the predictive uncertainty of all test subjects, can be gradually eliminated with sufficient information for learning the pairwise association. The results in Table 1 and Table 3 show that MCED uncertainty estimation, on one hand, can be adopted as an ensemble approach to improve the diagnostic performance (+1.62% F1-score), and on the other hand, can be used to detect patients with highly uncertain diagnostic value, which is important for safety-critical CAD systems in reducing misdiagnosis rate.

### 3.2 Alzheimer’s Disease Prediction on ADNI

**ADNI** [1] is a large-scale database and contains longitudinal brain MRI, PET data, genetic, and phenotypic information of over 1700 adults for Alzheimer’s

disease (AD) study. In this work, we select the same 1675 samples with Mild Cognitive Impairments (MCI) used in Parisot [17] to facilitate a fair comparison, among which 843 acquisitions will convert to AD as diagnosed during the follow-up. For the imaging feature vector, we use the volumes of  $C = 138$  segmented brain structures extracted from MRI using MALP-EM [12], which are proven effective biomarkers for AD assessment [18]. We use phenotypic (age, gender) and genetic (APOE [16]) data for computing the pairwise association.

**TADPOLE.** [15] is a preprocessed subset of ADNI, consisting of 557 subjects each with over 350 multimodal features. Following AIG [9], the task is to classify each subject into three classes: cognitive normal, MCI, and AD. We use the segmentation features derived from MRI and PET data to obtain a 340-dimensional feature vector for each subject and use the phenotypic data, APOE and FDG-PET biomarkers for the graph construction.



**Fig. 2.** Comparative boxplots on ADNI (a, b) and TADPOLE (c) for Alzheimer’s disease prediction. Results are computed from 10-fold cross-validation.

**Results.** Comparative boxplots for ten folds between the four methods are shown in Fig. 2. We can observe that the proposed method (EV-GCN+MCED) outperforms the competing methods on both datasets. For prediction AD conversion on ADNI, we achieve an average accuracy of 79.4%, corresponding to a 3.9% increase over the competing method Parisot [17]. For TADPOLE, our method obtains an average accuracy of 87.8%, outperforming the recent SoTA method AIG [9] (84.3%). The results also imply the generalizability of our method.

### 3.3 Ocular Disease Diagnosis on the ODIR Dataset

**Dataset and Experimental Setting.** The ODIR dataset [2] shares fundus photographs and non-imaging data including age, gender and diagnostic words of 3000 patients. Each patient has 8 binary labels for 7 types of ocular diseases including diabetes, glaucoma, etc. We compare our method with two recent SoTA CNNs [21, 22] and the graph-based SoTA AIG [9], using 5-fold cross-validation. For population graph construction, we use the compared CNN without classification layers as a feature extractor to derive a  $C$  dimensional feature vector



from the fundus images (for both eyes) of a patient ( $C = 3072$  for InceptionV4 [21] and  $C = 2048$  for EfficientNet-B0 [22]), and use the demographic data for association modeling. Diagnostic words are not used to avoid label leaking.

**Table 4.** Quantitative results on ODIR. (I) or (E): InceptionV4 [21] or EfficientNet [22] is used for imaging feature extraction. D: Diabetes, G: Glaucoma, C: Cataract, Overall: all 8 classes.

Methods	D	G	C	Overall
InceptionV4 [21]	64.26	69.89	95.34	$84.00 \pm 11.35$
EV-GCN (I)	68.49	70.31	94.22	$86.62 \pm 11.26$
EfficientNet [22]	66.90	71.91	<b>95.91</b>	$84.31 \pm 12.11$
EV-GCN (E)	<b>70.78</b>	<b>73.24</b>	95.18	<b><math>87.63 \pm 9.88</math></b>
AIG [9] (E)	58.61	68.08	87.60	$78.47 \pm 14.79$

Table 4 shows the comparative results in terms of AUC for different types of ocular diseases. We see that the EV-GCN can improve the classification performance for both EfficientNet and InceptionV4 (i.e., InceptionResNet-V2 [21]) on ODIR, e.g. 4% improvement for Diabetic detection. On average, the proposed method can improve the performance of a pre-trained SoTA CNN by 2.97% for fundus image classification, by learning to incorporate the complementary non-imaging data encoded in the graph. It is interesting to note that the static graph-based method [9], where the required thresholds are already finetuned, degrades the performance for EfficientNet, which reassures the robustness of constructing a learnable population graph compared to a hand-crafted one.

## 4 Discussion and Conclusions

In this paper, we have proposed a generalizable graph-convolutional framework to tackle the challenges in learning from multi-modal data for disease prediction. Unlike previous methods, the proposed method does not hand-engineer a similarity population graph but learn to construct the graph connectivity which is mathematically proven to be optimizable with GCNs. The proposed Monte-Carlo edge dropout is the first study on graph uncertainty estimation for GCNs and is experimentally validated to be beneficial, while we admit that it requires further theoretical justification in future work. Extensive experimental results show that the proposed method can achieve superior performance on brain analysis and ocular disease prediction. Additionally, the estimated predictive uncertainty allows detecting the uncertain samples for clinical intervention, contributing to a safer deep learning-assisted diagnosis system. We believe such an extendable method can have a great impact in unlocking a better use of multi-modal data in populations for computer-aided diagnosis in clinics.

## References

1. Alzheimer's Disease Neuroimaging Initiative (2019). <http://adni.loni.usc.edu/>
2. Ocular Disease Intelligent Recognition (2019). <https://odir2019.grand-challenge.org/dataset/>
3. Abraham, A., et al.: Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. *NeuroImage* **147**, 736–745 (2017)
4. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: *Advances in Neural Information Processing Systems*, pp. 3844–3852 (2016)
5. Di Martino, A., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **19**(6), 659–667 (2014)
6. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: *International Conference on Machine Learning*, pp. 1050–1059 (2016)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
8. Huang, Y., Chung, A.C.S.: Evidence localization for pathology images using weakly supervised learning. In: Shen, D., et al. (eds.) *MICCAI 2019*. LNCS, vol. 11764, pp. 613–621. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32239-7\\_68](https://doi.org/10.1007/978-3-030-32239-7_68)
9. Kazi, A., et al.: InceptionGCN: receptive field aware graph convolutional network for disease prediction. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) *IPMI 2019*. LNCS, vol. 11492, pp. 73–85. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-20351-1\\_6](https://doi.org/10.1007/978-3-030-20351-1_6)
10. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: *Advances in Neural Information Processing Systems*, pp. 5574–5584 (2017)
11. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016)
12. Ledig, C., Heckemann, R.A., Rueckert, D., et al.: Robust whole-brain segmentation: application to traumatic brain injury. *Med. Image Anal.* **21**(1), 40–58 (2015)
13. Li, Q., Han, Z., Wu, X.M.: Deeper insights into graph convolutional networks for semi-supervised learning. In: *Thirty-Second AAAI Conference on Artificial Intelligence* (2018)
14. Litjens, G., Kooi, T., Sánchez, C.I., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
15. Marinescu, R.V., Oxtoby, N.P., Alexander, D.C., et al.: Tadpole challenge: prediction of longitudinal evolution in Alzheimer's disease. *arXiv preprint arXiv:1805.03909* (2018)
16. Mosconi, L., Perani, D., Padovani, A., et al.: MCI conversion to dementia and the APOE genotype: a prediction study with FDG-PET. *Neurology* **63**(12), 2332–2340 (2004)
17. Parisot, S., et al.: Spectral graph convolutions for population-based disease prediction. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017*. LNCS, vol. 10435, pp. 177–185. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-66179-7\\_21](https://doi.org/10.1007/978-3-319-66179-7_21)
18. Ries, M.L., et al.: Magnetic resonance imaging characterization of brain structure and function in mild cognitive impairment: a review. *J. Am. Geriatr. Soc.* **56**(5), 920–934 (2008)

19. Rong, Y., Huang, W., Xu, T., Huang, J.: The truly deep graph convolutional networks for node classification. arXiv preprint [arXiv:1907.10903](https://arxiv.org/abs/1907.10903) (2019)
20. Shuman, D.I., Narang, S.K., Vandergheynst, P., et al.: The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* **30**(3), 83–98 (2013)
21. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-First AAAI Conference on Artificial Intelligence* (2017)
22. Tan, M., Le, Q.: Efficientnet: rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*, pp. 6105–6114 (2019)
23. Trojanowski, J.Q., et al.: Update on the biomarker core of the Alzheimer’s disease neuroimaging initiative subjects. *Alzheimer’s Dement.* **6**(3), 230–238 (2010)
24. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. arXiv preprint [arXiv:1901.00596](https://arxiv.org/abs/1901.00596) (2019)
25. Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K.I., Jegelka, S.: Representation learning on graphs with jumping knowledge networks. In: *International Conference on Machine Learning*, pp. 5453–5462 (2018)
26. Xu, T., Zhang, H., Huang, X., Zhang, S., Metaxas, D.N.: Multimodal deep learning for cervical dysplasia diagnosis. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9901, pp. 115–123. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46723-8\\_14](https://doi.org/10.1007/978-3-319-46723-8_14)
27. Yu, L., Wang, S., Li, X., Fu, C.-W., Heng, P.-A.: Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: Shen, D., et al. (eds.) *MICCAI 2019*. LNCS, vol. 11765, pp. 605–613. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32245-8\\_67](https://doi.org/10.1007/978-3-030-32245-8_67)