# Multimodal Brain Age Estimation Using Interpretable Adaptive Population-Graph Learning

Kyriaki-Margarita Bintsi[1]([✉]), Vasileios Baltatzis[1,2],
Rolandos Alexandros Potamias[1], Alexander Hammers[2], and Daniel Rueckert[1,3]

[1] Department of Computing, Imperial College London, London, UK
m.bintsi19@imperial.ac.uk
[2] Biomedical Engineering and Imaging Sciences, King's College London, London, UK
[3] Technical University of Munich, Munich, Germany

**Abstract.** Brain age estimation is clinically important as it can provide valuable information in the context of neurodegenerative diseases such as Alzheimer's. Population graphs, which include multimodal imaging information of the subjects along with the relationships among the population, have been used in literature along with Graph Convolutional Networks (GCNs) and have proved beneficial for a variety of medical imaging tasks. A population graph is usually static and constructed manually using non-imaging information. However, graph construction is not a trivial task and might significantly affect the performance of the GCN, which is inherently very sensitive to the graph structure. In this work, we propose a framework that learns a population graph structure optimized for the downstream task. An attention mechanism assigns weights to a set of imaging and non-imaging features (phenotypes), which are then used for edge extraction. The resulting graph is used to train the GCN. The entire pipeline can be trained end-to-end. Additionally, by visualizing the attention weights that were the most important for the graph construction, we increase the interpretability of the graph. We use the UK Biobank, which provides a large variety of neuroimaging and non-imaging phenotypes, to evaluate our method on brain age regression and classification. The proposed method outperforms competing static graph approaches and other state-of-the-art adaptive methods. We further show that the assigned attention scores indicate that there are both imaging and non-imaging phenotypes that are informative for brain age estimation and are in agreement with the relevant literature.

**Keywords:** Brain age regression · Interpretability · Graph Convolutional Networks · Adaptive graph learning

# 1   Introduction

Healthy brain aging follows specific patterns [2]. However, various neurological diseases, such as Alzheimer's disease [8], Parkinson's disease [23], and schizophrenia [18], are accompanied by an abnormal accelerated aging of the human brain. Thus, the difference between the biological brain age of a person and their chronological age can show the deviation from the healthy aging trajectory, and may prove to be an important biomarker for neurodegenerative diseases [6,9].

Recently, graph-based methods have been explored for brain age estimation as graphs can inherently combine multimodal information by integrating the subjects' neuroimaging information as node features and, through a similarity metric, the associations among subjects through as edges that connect these nodes [21]. However, in medical applications, the construction of a population-graph is not always simple as there are various ways subjects could be considered similar.

GCNs [16] have been extensively [1] used in the medical domain for node classification tasks, such as Alzheimer's prediction [15,21] and Autism classification [4]. They take graphs as input and, in most cases, the graph structure is predefined and static. GCN performance is highly dependent on the graph structure. This has been correlated in related literature with the heterophily of graphs in the semi-supervised node classification tasks, which refers to the case when the nodes of a graph are connected to nodes that have dissimilar features and different class labels [29]. It has been shown that if the homophily ratio is very low, a simple Multi-Layer Perceptron (MLP) that completely ignores the structure, can outperform a GCN [30].

A way to address this problem is through adaptive graph learning [27], which learns the graph structure through training. In the medical domain, there is little ongoing research on the topic [13,28]. However, the adaptive graphs in [7,13,14]
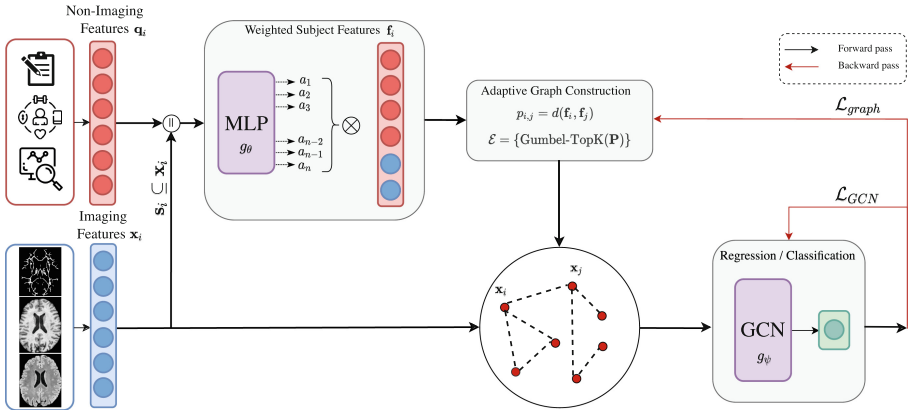


**Fig. 1.** Proposed methodology: The non-imaging features $\mathbf{q}_i$ and a subset of the imaging features $\mathbf{s}_i \subseteq \mathbf{x}_i$ per subject $i$ are used as input to a MLP, which produces attention weights for each one of these features. Based on these, the edges of the graph are stochastically sampled using the Gumbel-Top-k trick. The constructed graph, which uses the imaging features $\mathbf{X}$ as node features is used to train a GCN for brain age estimation tasks.

are connected based on the imaging features and do not take advantage of the associations of the non-imaging information for the edges. In [11], non-imaging information is used for graph connectivity, but the edges are already pre-pruned similar to [21], and only the weights of the edges can be modified. In [26], even though the graph is dynamic, it is not being learnt. While the extracted graphs in these studies are optimized for the task at hand, there is no explanation for the node connections that were proposed. Given that interpretability is essential in the medical domain [24], since the tools need to be trusted by the clinicians, an adaptive graph learnt during the training, whose connections are also interpretable and clear can prove important. Additionally, most existing works focus on brain age classification in four bins, usually classifying the subjects per decade [7,13,14]. Age regression, which is a more challenging task, has not been extensively explored with published results not reaching sufficient levels of performance [25].

**Contributions.** This paper has the following contributions: 1) We combine imaging and non-imaging information in an attention-based framework to learn adaptively an optimized graph structure for brain age estimation. 2) We propose a novel graph loss that enables end-to-end training for the task of brain age regression. 3) Our framework is inherently interpretable as the attention mechanism allows us to rank all imaging and non-imaging phenotypes according to their significance for the task. 4) We evaluate our method on the UK Biobank (UKBB) and achieve state-of-the-art results on the tasks of brain age regression and classification. The code can be found on GitHub at: https://github.com/bintsi/adaptive-graph-learning.

## 2    Methods

Given a set of $N$ subjects with $M$ features $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N] \in \mathbb{R}^{N \times M}$ and labels $\mathbf{y} \in \mathbb{R}^N$, a population graph is defined as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is a set of nodes, one per subject, and $\mathcal{E}$ is a set of paired nodes that specifies the connectivity of the graph, meaning the edges of the graph. To create an optimized set of edges for the task of brain age estimation, we leverage a set of non-imaging phenotypes $\mathbf{q}_i \in \mathbb{R}^Q$ and a set of imaging phenotypes $\mathbf{s}_i \in \mathbb{R}^S$ per subject $i$ through an attention-based framework. The imaging phenotypes are a subset of the imaging features $\mathbf{s}_i \subseteq \mathbf{x}_i$. The phenotypes are selected according to [5]. The resulting graph is used as input in a GCN that performs node-level prediction tasks. Here, we give a detailed description of the proposed architecture, in which the connectivity of the graph $\mathcal{E}$, is learnt through end-to-end training in order to find the optimal graph structure for the task of brain age estimation. An outline of the proposed pipeline may be found in Fig. 1.

**Attention Weights Extraction.** Based on the assumption that not all phenotypes are equally important for the construction of the graph, we train a MLP $g_\theta$, with parameters $\theta$, which takes as input both the non-imaging features $\mathbf{q}_i$ and the imaging features $\mathbf{s}_i$ for every subject $i$ and outputs an attention weight

vector $\mathbf{a} \in \mathbb{R}^{Q+S}$, where every element of $\mathbf{a}$ corresponds to a specific phenotype. Intuitively, we expect that the features that are relevant to brain age estimation will get attention weights close to 1, and close to 0 otherwise. Since we are interested in the overall relevance of the phenotypes for the task, the output weights need to be global and apply to all of the nodes. To do so, we average the attention weights across subjects and normalize them between 0 and 1.

**Edge Extraction.** The weighted phenotypes for each subject $i$ are calculated as $\mathbf{f}_i = \mathbf{a} \odot (\mathbf{q}_i \,\|\, \mathbf{s}_i)$, where $\mathbf{f}_i \in \mathbb{R}^{Q+S}$, $\mathbf{a}$ are the attention weights produced by the MLP $g_\theta$, $(\cdot \,\|\, \cdot)$ denotes the concatenation function between two vectors and $\odot$ denotes the Hadamard product. We further define the probability $p_{ij}(\mathbf{f}_i, \mathbf{f}_j; \theta, t)$ of a pair of nodes $(i, j) \in \mathcal{V}$ to be connected in Eq. (1):

$$p_{ij}(\mathbf{f}_i, \mathbf{f}_j; \theta, t) = e^{-td(\mathbf{f}_i, \mathbf{f}_j)^2} \tag{1}$$

where $t$ is a learnable parameter and $d$ is a distance function that calculates the distance between the weighted phenotypes of two nodes. To keep the memory cost low, we create a sparse k-degree graph. We use the Gumbel-Top-k trick [17], which acts as a stochastic relaxation of the kNN rule, in order to sample k edges for every node according to the probability matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$. Since there is stochasticity in the sampling scheme, multiple runs are performed at inference time and the predictions are averaged.

**Optimization.** The extracted graph is used as input, along with the imaging features $\mathbf{X}$, which are used as node features, to a GCN $g_\psi$, with parameters $\psi$, which comprises of a number of graph convolutional layers, followed by fully connected layers. The pipeline is trained end-to-end with a loss function $\mathcal{L}$ that consists of two components and is defined as in Eq. (2).

$$\mathcal{L} = \mathcal{L}_{GCN} + \mathcal{L}_{graph}. \tag{2}$$

The first component, $\mathcal{L}_{GCN}$, is optimizing the GCN, $g_\psi$. For regression we use the Huber loss [12], while for classification we use the Cross Entropy loss function. The second component, $\mathcal{L}_{graph}$, optimizes the MLP $g_\theta$, whose output are the phenotypes' attention weights. However, the graph is sparse with discrete edges and hence the network cannot be trained with backpropagation as is. To alleviate this issue, we formulate our graph loss in a way that rewards edges that lead to correct predictions and penalize edges that lead to wrong predictions. Inspired by [13], where a similar approach was used for classification, the proposed graph loss function is designed for regression instead and is defined in Eq. (3):

$$\mathcal{L}_{graph} = \sum_{(i,j) \in \mathcal{E}} \rho(y_i, g_\psi(\mathbf{x}_i)) \, log(p_{ij}), \tag{3}$$

where $\rho(\cdot, \cdot)$ is the reward function which is defined in Eq. (4):

$$\rho(y_i, g_\psi(\mathbf{x}_i)) = |y_i - g_\psi(\mathbf{x}_i)| - \varepsilon \tag{4}$$

Here $\varepsilon$ is the null model's prediction (i.e. the average brain age of the training set). Intuitively, when the prediction error $|y_i - g_\psi(\mathbf{x}_i)|$ is smaller than the null

**Table 1.** Performance of the proposed approach in comparison with a traditional machine learning model, static graph baselines, and the state-of-the-art for age regression (left), and 4-class age classification (right). (MAE in years)

| Method | MAE | $r$ score |
|---|---|---|
| Linear Regression | 3.82 | 0.75 |
| Static (Node features) | 3.87 | 0.75 |
| Static (Phenotypes) | 3.98 | 0.74 |
| DGM | 3.72 | 0.75 |
| **Proposed** | **3.61** | **0.79** |

| Method | Accuracy | AUC | F1 |
|---|---|---|---|
| Logistic Regression | 0.54 | 0.79 | 0.54 |
| Static (Node features) | 0.53 | 0.75 | 0.52 |
| Static (Phenotypes) | 0.52 | 0.75 | 0.51 |
| DGM | 0.55 | **0.80** | 0.55 |
| **Proposed** | **0.58** | **0.80** | **0.56** |

**Table 2.** Ablation studies. Left: Effect of the choice of phenotypes used for the connection of the edges. Right: Performance of the proposed using different distances on the phenotypes to connect the edges. (MAE in years)

| No. of phenotypes | MAE | $r$ score |
|---|---|---|
| 20 (non-imaging only) | 4.63 | 0.68 |
| **35 (non-imaging & imaging)** | **3.61** | **0.79** |
| 50 (non-imaging & imaging) | 3.65 | 0.78 |
| 68 (imaging only) | 3.73 | 0.77 |

| Similarity Metric | MAE | $r$ score |
|---|---|---|
| Random | 5.59 | 0.61 |
| **Euclidean** | **3.61** | **0.79** |
| Cosine | 3.7 | 0.78 |
| Hyperbolic | 4.08 | 0.72 |

model's prediction then the reward function will be negative. In turn, this will encourage the maximization of $p_{ij}$ so that $\mathcal{L}_{graph}$ is minimized.

## 3   Experiments

**Dataset.** The proposed framework is evaluated on the UKBB [3], which provides not only a wide collection of images of vital organs, including brain scans, but also various non-imaging information, such as demographics, and biomedical, lifestyle, and cognitive performance measurements. Hence, it is perfectly suitable for brain age estimation tasks that incorporate the integration of imaging and non-imaging information. Here, we use 68 neuroimaging phenotypes and 20 non-imaging phenotypes proposed by [5] as the ones most relevant to brain age in UKBB. The neuroimaging features are provided by UKBB, and include measurements extracted from structural MRI and diffusion weighted MRI. All phenotypes are normalized from 0 to 1. The age range of the subjects is 47–81 years. We only keep the subjects that have available the necessary phenotypes ending up with about 6500 subjects. We split the dataset into 75% for the training set, 5% for the validation set, and 20% for the test set. Our pipeline has been primarily designed to tackle the challenging regression task and therefore the main experiment is brain age regression. We also evaluate our framework on the 4-class classification task that has been used in other related papers.

**Baselines.** Given that a GCN trained on a meaningless graph can perform even worse than a simple regressor/classifier, our first baseline is a Linear/Logistic
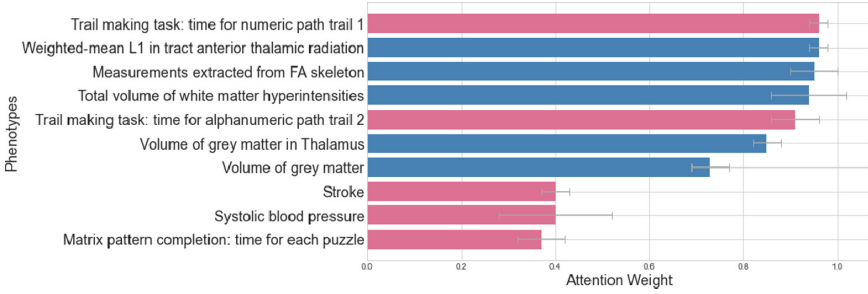
**Fig. 2.** Attention weights for the top 5 imaging (blue) and top 5 non-imaging phenotypes (pink) averaged across 10 runs of the pipeline. (Color figure online)

Regression model. For the second baseline, we construct a static graph based on a similarity metric, more specifically cosine similarity, of a set of features (either node features or non-imaging and imaging phenotypes) using the kNN rule with $k = 5$ and train a GCN on that graph. Using euclidean distance as the similarity metric leads to worse performance and is therefore not explored for the baselines. We also compare our method with DGM, which is the state-of-the-art on graph learning for medical applications [13]. Since this work is only applicable for classification tasks, we extend it to regression, implementing the graph loss function we used for our pipeline as well.

**Implementation Details.** The GCN architecture uses ReLU activations and consists of one graph convolutional layer with 512 units and one fully connected layer with 128 units before the regression/classification layer. The number and the dimensions of the layers are determined through hyperparameter search based on validation performance. The networks are trained with the AdamW optimizer [20], with a learning rate of 0.005 for 300 epochs and utilize an early stopping scheme. The average brain age of the training set, which is used in Eq. (3), is $\varepsilon = 6$. We use PyTorch [22] and a Titan RTX GPU. The reported results are computed by averaging 10 runs with different initializations.

### 3.1   Results

**Regression.** For the evaluation of the pipeline in the regression task, we use Mean Absolute Error (MAE), and the Pearson Correlation Coefficient ($r$ score). A summary of the performance of the proposed and competing methods is available at Table 1. Linear regression (MAE = 3.82) outperforms GCNs trained on static graphs whether these are based on the node features (MAE = 3.87) or leverage the phenotypes (MAE = 3.98). The DGM outperforms the baselines (MAE = 3.72). The proposed method outperforms all others (MAE = 3.61).

**Classification.** For the classification task, we divide the subjects into four balanced classes. The metrics used for the classification task to evaluate the performance of the model are accuracy, the area under the ROC curve (AUC), which is
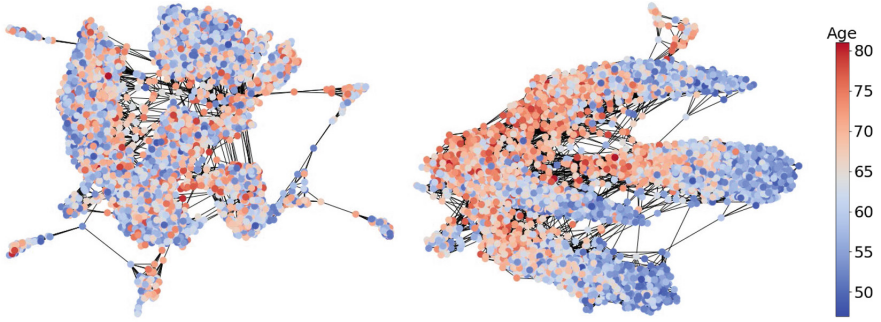
**Fig. 3.** Population-graph visualizations. Left: graph constructed based on the cosine similarity of the phenotypes. Right: graph constructed based on the cosine similarity of the weighted phenotypes, where the attention weights are the ones extracted from the proposed pipeline. The color of the nodes corresponds to the subject's age. It is evident that using the attention weights increases the homophily of the graph.

defined as the average of the AUC curve of every class, and the Macro F1-score (Table 1). A similar trend to the regression task appears here, with the proposed method reaching top performance with 58% accuracy.

Our hypothesis that the construction of a pre-defined graph structure is suboptimal, and might even hurt performance, is confirmed. Both for the classification and regression tasks, GCNs trained on static graphs do not outperform a simple linear model that ignores completely the structure of the graph. The proposed approach proves that not all phenotypes are equally important for the construction of the graph, and that giving attention weights accordingly increases the performance for both tasks.

### 3.2   Ablation Studies

**Number of Phenotypes.** We perform an ablation test to investigate the effect of the number of features used for the extraction of the edges. We used only non-imaging phenotypes, only imaging phenotypes, or a combination of both (Table 2). The combination of imaging and non-imaging phenotypes (MAE = 3.61) performs better than using either one of them, while adding more imaging phenotypes does not necessarily improve performance.

**Distance Metrics.** Moreover, we explore how different distance metrics affect performance. Here, euclidean, cosine, and hyperbolic [19] distances are explored. We also include a random edge selection to examine whether using the phenotypic information improves the results. The results (Table 2) indicate that euclidean distance performs the best, closely followed by cosine similarity. Hyperbolic performed a bit worse (MAE = 4 years), while random edges gave a MAE of 5.59. Regardless of the distance metric, phenotypes do incorporate valuable information as performance is considerably better than using random edges.

### 3.3   Interpretability

A very important advantage of the pipeline is that the graph extracted through the training is interpretable, in terms of why two nodes are connected or not. Visualizing the attention weights given to the phenotypes, we get an understanding of the features that are the most relevant to brain aging. The regression problem is clinically more important, thus we will focus on this in this section. A similar trend was presented for the classification task as well.

The imaging and non-imaging phenotypes that were given the highest attention scores in the construction of the graph can be seen in Fig. 2. We color the non-imaging phenotypes in pink, and the imaging ones in blue. A detailed list of the names of the phenotypes along with the attention weights given by the pipeline can be found in the Supplementary Material. The non-imaging phenotypes that were given the highest attention weights, were two cognitive tasks, the numeric and the alphanumeric trail making tasks. Systolic blood pressure, and stroke, were the next most important non-imaging phenotypes, even though they were not as important as some of the imaging features. Various neuroimaging phenotypes were considered important, such as information regarding the tract anterior thalamic radiaton, the volume of white matter hyperintensities, gray matter volumes, as well as measurements extracted from the FA (fractional anisotropy, a measure of the connectivity of the brain) skeleton. Our findings are in agreement with the relevant literature [5].

Apart from the attention weights, we also visualize the population graph that was used as the static graph of the baseline (Fig. 3(left)). This population graph is constructed based on the cosine similarity of the phenotypes, with all the phenotypes playing an equally important role for the connectivity. In addition, we visualize the population graph using the cosine similarity of the weighted phenotypes (Fig. 3(right)), with the attention weights provided by our trained pipeline. In both of the graph visualizations, the color of each node corresponds to the subject's age. It is clear that learning the graph through our pipeline results in a population graph where subjects with similar ages end up in more compact clusters, whereas the static graph does not demonstrate any form of organization. Since GCNs are affected by a graph construction with low homophily, it is reasonable that the static graphs perform worse than a simple machine learning method, and why the proposed approach manages to produce state-of-the-art results. Using a different GNN architecture that is not as dependent on the constructed graph's homophily could prove beneficial [10,29].

## 4   Conclusion

In this paper, we propose an end-to-end pipeline for adaptive population-graph learning that is optimized for brain age estimation. We integrate multimodal information and use attention scores for the construction of the graph, while also increasing interpretability. The graph is sparse, which minimizes the computational costs. We implement the approach both for node regression and node classification and show that we outperform both the static graph-based and

state-of-the-art approaches. We finally provide an insight into the most important phenotypes for graph construction, which are in agreement with related neurobiological literature.

In future work, we plan to extract node features from the latent space of a CNN. Training end-to-end will focus on latent imaging features that are also important for the GCN. Such features would potentially be more expressive and improve the overall performance. Finally, we leverage the UKBB because of the wide variety of multimodal data it contains, which makes it a perfect fit for brain age estimation but as a next step we also plan to evaluate our framework on different tasks and datasets.

# References

1. Ahmedt-Aristizabal, D., Armin, M.A., Denman, S., Fookes, C., Petersson, L.: Graph-based deep learning for medical diagnosis and analysis: past, present and future. Sensors **21**(14), 4758 (2021)
2. Alam, S.B., Nakano, R., Kamiura, N., Kobashi, S.: Morphological changes of aging brain structure in MRI analysis. In: 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS), pp. 683–687. IEEE (2014)
3. Alfaro-Almagro, F., et al.: Image processing and quality control for the first 10,000 brain imaging datasets from UK biobank. NeuroImage **166**, 400–424 (2018)
4. Anirudh, R., Thiagarajan, J.J.: Bootstrapping graph convolutional neural networks for autism spectrum disorder classification. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3197–3201. IEEE (2019)
5. Cole, J.H.: Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. Neurobiol. Aging **92**, 34–42 (2020)
6. Cole, J.H., et al.: Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. Neuroimage **163**, 115–124 (2017)
7. Cosmo, L., Kazi, A., Ahmadi, S.-A., Navab, N., Bronstein, M.: Latent-graph learning for disease prediction. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 643–653. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_62
8. Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q.: Prediction of MCI to ad conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol. Aging **32**(12), 2322-e19 (2011)
9. Franke, K., Gaser, C.: Ten years of brainage as a neuroimaging biomarker of brain aging: what insights have we gained? Front. Neurol. 789 (2019)
10. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
11. Huang, Y., Chung, A.C.S.: Edge-variational graph convolutional networks for uncertainty-aware disease prediction. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 562–572. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_55

12. Huber, P.J.: Robust estimation of a location parameter. In: Kotz, S., Johnson, N.L. (eds.) Breakthroughs in Statistics: Methodology and Distribution, pp. 492–518. Springer, New York (1992). https://doi.org/10.1007/978-1-4612-4380-9_35

13. Kazi, A., Cosmo, L., Ahmadi, S.A., Navab, N., Bronstein, M.: Differentiable graph module (DGM) for graph convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **45**(2), 1606–1617 (2022)

14. Kazi, A., Farghadani, S., Navab, N.: IA-GCN: interpretable attention based graph convolutional network for disease prediction. arXiv preprint arXiv:2103.15587 (2021)

15. 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

16. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)

17. Kool, W., Van Hoof, H., Welling, M.: Stochastic beams and where to find them: the gumbel-top-k trick for sampling sequences without replacement. In: International Conference on Machine Learning, pp. 3499–3508. PMLR (2019)

18. Koutsouleris, N., et al.: Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders. Schizophr. Bull. **40**(5), 1140–1153 (2014)

19. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., Boguná, M.: Hyperbolic geometry of complex networks. Phys. Rev. E **82**(3), 036106 (2010)

20. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)

21. Parisot, S., et al.: Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. Med. Image Anal. **48**, 117–130 (2018)

22. Paszke, A., et al.: PyTorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, vol. 32 (2019)

23. Reeve, A., Simcox, E., Turnbull, D.: Ageing and Parkinson's disease: why is advancing age the biggest risk factor? Ageing Res. Rev. **14**, 19–30 (2014)

24. Shaban-Nejad, A., Michalowski, M., Buckeridge, D.L.: Explainability and interpretability: keys to deep medicine. In: Shaban-Nejad, A., Michalowski, M., Buckeridge, D.L. (eds.) Explainable AI in Healthcare and Medicine. SCI, vol. 914, pp. 1–10. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-53352-6_1

25. Stankeviciute, K., Azevedo, T., Campbell, A., Bethlehem, R., Lio, P.: Population graph GNNs for brain age prediction. In: ICML Workshop on Graph Representation Learning and Beyond (GRL+), pp. 17–83 (2020)

26. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. (tog) **38**(5), 1–12 (2019)

27. Wei, S., Zhao, Y.: Graph learning: a comprehensive survey and future directions. arXiv preprint arXiv:2212.08966 (2022)

28. Zheng, S., et al.: Multi-modal graph learning for disease prediction. IEEE Trans. Med. Imaging **41**(9), 2207–2216 (2022)

29. Zheng, X., Liu, Y., Pan, S., Zhang, M., Jin, D., Yu, P.S.: Graph neural networks for graphs with heterophily: a survey. arXiv preprint arXiv:2202.07082 (2022)

30. Zhu, J., Yan, Y., Zhao, L., Heimann, M., Akoglu, L., Koutra, D.: Beyond homophily in graph neural networks: current limitations and effective designs. Adv. Neural. Inf. Process. Syst. **33**, 7793–7804 (2020)