Edge-Aware Hierarchical Graph Transformer to Decode Brain Arterial Network

Kaiyu Zhang[1 [0000−0001−7208−3114]], Li Chen[2 [0000−0003−0233−4576]], Wenjin Liu[3], Taewon Kim[4 [0000−0001−7071−1455]], Xin Wang[2 [0000−0001−9546−6038]], Yin Guo[1 [0000−0002−6006−9849]], Zhiwei Tan[2 [0000−0002−7964−0283]], Zhensen Chen[5 [0000−0001−7975−2528]], Angie Tang[6], Xihai Zhao[7 [0000−0001−8953−0566]], Thomas S. Hatsukami[8 [0000−0002−0129−5846]], Mahmud Mossa-Basha[9 [0000−0001−7798−8158]], Niranjan Balu[9 [0000−0002−1489−6683]], and Chun Yuan[9 [0000−0002−6154−5106]]

[1] Department of Biomedical Engineering, University of Washington, Seattle, WA, USA
[2] Department of Electrical and Computer Engineering, University of Washington, Seattle, Washington, USA
[3] Department of Nephrology, Clinical Medical College, Northern Jiangsu People's Hospital, Yangzhou University, Yangzhou, China
[4] Department of Neurology, Incheon St. Mary's Hospital, The Catholic University of Korea, Korea
[5] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China
[6] University of Washington, Seattle, WA, USA
[7] Center for Biomedical Imaging Research, Department of Biomedical Engineering, Tsinghua University, Beijing, China
[8] Department of Surgery, University of Washington, Seattle, Washington, USA
[9] Department of Radiology, University of Washington, Seattle, Washington, USA
cyuan@uw.edu

Abstract. Many biomedical data exhibit intrinsic graph-like properties, making graph neural networks (GNNs) widely adopted modeling tools. The brain arterial network (BAN) represents the most complex arterial network in humans, where conventional GNNs struggle to capture critical long-range relationships. Recent graph transformers have enabled modeling of these long-range dependencies through attention mechanisms; however, they face challenges in incorporating hierarchical information, especially when strong anatomical priors exist within the graph structure. While some approaches have attempted to integrate hierarchical information into graph transformers, they primarily focus on node feature aggregation, despite BAN's most clinically significant features residing in edges rather than nodes. To address these limitations, we propose a hierarchical graph transformer (HGT) with edge-aware structural encoding that better incorporates anatomical and multi-scale structural information. Our approach achieves state-of-the-art performance across all 11 tasks. This work lays the foundation for individualized risk assessment that complements traditional systemic risk evaluation methods.

Keywords: Brain Arterial Network · Graph Transformer · Cardiovascular Risk Prediction.

1 Background

Cardiovascular disease remains the leading cause of death worldwide. Prevention strategies predominantly rely on the Framingham Risk Score (FRS) to estimate 10-year cardiovascular risk using systemic metrics such as blood pressure, cholesterol, age, and smoking status [6]. However, the FRS evaluates only these external factors while overlooking internal vascular imaging biomarkers that directly reflect vascular health. Although magnetic resonance angiography (MRA) is widely employed to assess localized vascular pathology (e.g., stenosis), its application in analyzing global vasculature structure remains limited. In particular, the brain arterial network (BAN) represents one of the most complex vascular systems in the human body; numerous studies have demonstrated its association with aging, hypertension, and cognitive decline [23,13,4], yet research progress is impeded by the network's multi-hierarchical structure and the significant challenges in extracting full-brain BAN from MRA.

Recent research has demonstrated that human vasculature modeled by graph neural networks (GNNs) can effectively capture disease-related patterns from vascular graphs [22,18]. However, previous studies utilizing conventional GNN architectures, such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), have primarily focused on relatively small graphs[17,10], where their local message-passing operations prove insufficient for modeling larger, complex networks like the BAN. Recent advancements in graph transformers have addressed some of these limitations by leveraging global self-attention mechanisms [12,20,5]; nevertheless, standard graph transformers lack the explicit capability to model hierarchical patterns due to inductive bias, particularly when the graph exhibits significant anatomical structure. Furthermore, while several graph transformer variants have been proposed to incorporate hierarchical information[14], they typically emphasize node-feature aggregation, whereas in BAN, the edge features—representing critical arterial attributes such as radius, tortuosity, and length—are of paramount importance.

To address these methodological limitations in capturing the hierarchical complexity of brain arterial networks, we propose two contributions:

1. We propose a novel hierarchical graph transformer (HGT) with edge-aware structural encoding that leverages anatomical priors to capture both local and global vascular patterns within BAN.
2. Our proposed method achieves state-of-the-art performance across 11 predictive tasks, demonstrating the robust association between BAN and cardiovascular risk factors.

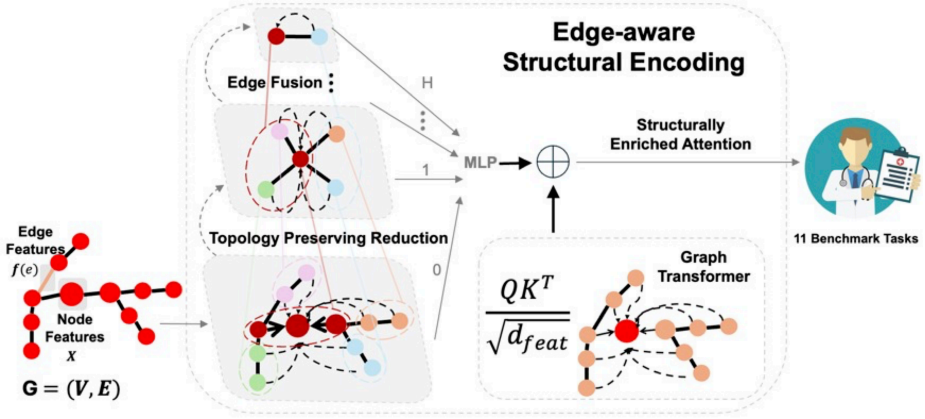

Fig. 1: Architecture of the proposed hierarchical graph transformer (HGT) with edge-aware structural encoding. The framework processes a vascular graph through hierarchical structural encoding pathways, transforming edge features into multi-level representations. These encodings are integrated with standard transformer attention, producing structurally enriched attention for capturing hierarchical information of the graph.

2 Methods

2.1 HGT with Edge-Aware Structural Encoding

We first propose a flexible module that can be appended to standard graph transformer architectures [20] to incorporate prior structural information from edge attributes, as shown in Fig.1. In our approach, we focus exclusively on edge features (e.g., artery length, radius, and tortuosity) and generate a multi-level structural encoding through two steps: (i) topology preserving reduction and (ii) multi-level coarsening via heavy-edge matching [1].

Topology Preserving Reduction. Given a vascular graph $G = (V, E)$ with node set V and edge set E, we first simplify G using a topology preserving reduction that retains all bifurcation and terminal nodes. For each edge $e \in E$, we define a feature vector $\mathbf{f}(e) = [\ell(e), r(e), \tau(e)]$, where $\ell(e)$ represents the vessel segment length, $r(e)$ the vessel radius, and $\tau(e)$ the vessel tortuosity. In our implementation, we apply a degree-2 node removal strategy to collapse linear segments. This yields a simplified graph $G^{(0)} = (V^{(0)}, E^{(0)})$ where each edge $e_{ij}^{(0)}$ represents an entire vessel segment between key nodes i and j. The new edge

attributes are aggregated as follows:

$$\begin{aligned} \ell_{ij}^{(0)} &= \sum_{e \in P(i,j)} \ell(e), \\ r_{ij}^{(0)} &= \frac{1}{|P(i,j)|} \sum_{e \in P(i,j)} r(e), \\ \tau_{ij}^{(0)} &= \frac{1}{|P(i,j)|} \sum_{e \in P(i,j)} \tau(e), \end{aligned} \tag{1}$$

where $P(i,j)$ denotes the set of edges along the original vessel segment connecting nodes i and j, and $|P(i,j)|$ represents the number of edges in this set. This reduction preserves the essential vascular branching structure while eliminating redundant nodes from continuous vessel segments.

Multi-Level Edge Fusion. To capture multi-scale structural information, we further coarsen $G^{(0)}$ via a heavy-edge matching (HEM) algorithm [1], generating a hierarchy of graphs

$$\{G^{(h)} = (V^{(h)}, E^{(h)})\}_{h=0}^{H},$$

where $V^{(h)}$ is the node set and $E^{(h)}$ is the edge set at level k. At each level k, an edge $e \in E^{(h)}$ with length $\ell(e)$ and radius $r(e)$ is assigned a weight based on a composite function:

$$m(e) = \frac{r(e)^2}{\ell(e)}, \tag{2}$$

which favors the merging of segments that are thick and short—a biologically meaningful heuristic for vascular networks. The HEM algorithm identifies a maximal matching of node pairs by greedily selecting edges with the highest weights, then contracts the matched edges to form the next coarser graph $G^{(h+1)}$. For any two nodes i, j in the original graph, we define the multi-level edge encoding as

$$\mathbf{M}_{ij} = \Big[m^{(0)}(i,j),\, m^{(1)}(i,j),\, \ldots,\, m^{(H)}(i,j) \Big] \in \mathbb{R}^{H+1}, \tag{3}$$

where $m^{(h)}(i,j)$ denotes the edge weight measure between the coarsened representations of nodes i and j in $G^{(h)}$, specifically $m^{(h)}(i,j) = m(e_{uv})$ where $u, v \in V^{(h)}$ are the respective mappings of nodes i, j after h levels of coarsening.

Structurally Enriched Attention. The multi-level edge encoding is then processed to generate an attention bias. For each node pair (i,j), we define

$$b_{ij} = \mathrm{MLP}\Big(\mathbf{M}_{ij}\Big), \tag{4}$$

where MLP is a trainable multi-layer perceptron that transforms the hierarchical edge weights into a scalar attention bias. Following the GPS framework [20], the graph representation matrix $X \in \mathbb{R}^{N \times d_{\text{feat}}}$ is first processed through a local message-passing layer that aggregates both node and edge information, where

N is the number of nodes and d_{feat} is the feature dimension. The transformer then computes

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V, \tag{5}$$

with $W_Q, W_K, W_V \in \mathbb{R}^{d_{\text{feat}} \times d_{\text{feat}}}$ as learnable parameters. The self-attention is then formulated as

$$\text{Attn}(X) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_{\text{feat}}}} + B\right)V, \tag{6}$$

where the bias matrix $B \in \mathbb{R}^{N \times N}$ is constructed by stacking the scalars b_{ij} for all node pairs. Adding the bias directly to attention scores allows the model to modulate attention based on multi-level structural information while preserving the content-based attention mechanism of the transformer.

Visualization of Structural Attention. We extract interpretable insights from our model by computing importance scores for nodes based on edge attention values. The edge importance scores are directly obtained from the attention coefficients of our graph attention layers. Node importance is then derived by accumulating the attention scores associated with each node's connected edges:

$$\text{NodeImp}(v_i) = \sum_{e_{ij} \in E(i)} \text{EdgeImp}(e_{ij}), \tag{7}$$

where $E(i)$ represents all edges connected to node i and $EdgeImp(e_{ij})$ represents the attention coefficient after the softmax function obtained from Equation (6). This approach allows us to visualize which nodes in the brain arterial network are most influential for the model's predictions.

2.2 Dataset Formation and Task Establishment

Cohort Description and Risk Factor Definitions. This cross-sectional analysis pooled baseline data from three East Asian cohorts [13,4]. A total of 402 participants were aged ≥ 45 years with no history of cardiovascular events. Blood pressure (SBP, DBP), Serum lipid profiles (HDL, LDL, total cholesterol), and smoking/diabetes status were obtained through standardized laboratory assays and questionnaires/medical records. The FRS was calculated to estimate 10-year cardiovascular risk (0–1) [6]. All studies adhered to IRB protocols. The distribution of cardiovascular risk factors across the cohort is illustrated in Fig.2(a).

Imaging Protocols and Graph Construction Process. 3D MRA images were acquired using Philips, GE, and Siemens scanners at 1.5/3.0 Tesla, with in-plane resolutions ranging from 0.35 mm to 0.5 mm and full brain coverage. 3D brain artery trace was constructed by [2]. Multiple reproducibility studies were conducted to validate the trace reconstruction accuracy across imaging protocols from various vendors and independent reviewer annotations [3,8,24]. Graphs were constructed using the `networkx` package [9]. Node features included 3D coordinates of artery trace, while edge features comprised artery radius, length, and tortuosity. If the original BAN was not fully connected, a virtual node

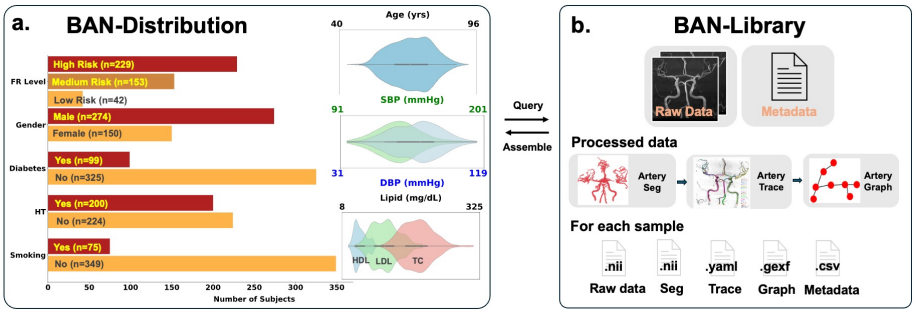

Fig. 2: Overview of the BAN dataset. (a) **BAN-Overview:** Demographic and clinical characteristics of the dataset, including FRS, gender, diabetes, hypertension (HT), smoking status, and distributions of age, blood pressure (SBP, DBP), and lipid levels (HDL, LDL, TC). (b) **BAN-Library:** A database containing raw MRA images, artery segmentation, traced artery centerlines, and graph representations, with corresponding metadata and extracted features.

was introduced to link disconnected components [11], with no additional node or edge features assigned to the virtual node. Finally, node and edge features were standardized per-subject to ensure consistency across the dataset. Detailed process was shown in Fig.2(b).

Prediction Tasks. Predictive tasks included **classification** of FRS levels (<0.1, 0.1–0.2, >0.2), hypertension, diabetes, and smoking status; and **regression** for FRS, age, SBP, DBP, HDL, LDL, and TC. For regression tasks, we report Mean Absolute Error (MAE) and the coefficient of determination (R^2), which quantifies the proportion of variance explained by the model. For classification tasks, we evaluate performance using micro-F1, macro-F1, and Area Under the ROC Curve (AUC); micro-F1 aggregates contributions from all classes, while macro-F1 averages F1 scores across classes to reflect performance on less frequent labels. Compared models included graph convolutional networks (GCN) [15], graph attention networks (GAT) [21], graph transformer (GT) [5], and GPS [20].

3 Experiments and Results

3.1 Experiments

Implementation Details. Experiments are conducted using a 5-fold cross-validation strategy for both classification and regression tasks. Hyperparameter tuning is performed via grid search managed by the Neural Network Intelligence tool [16] with a cosine learning rate scheduler. All models—including our proposed HGT and baseline models (GCN, GAT, GT, and GPS)—employ data augmentation [25] with a multiplier of 5 to ensure fairness and an adequate trade-off between training power and time. Our augmentation techniques add

Fig. 3: Visualization of node and edge attention importance in BAN for age prediction across three subjects of increasing age (48, 67, and 88 years). The intensity of red coloring indicates higher attention weights.

Gaussian noise (std=0.01) to node coordinates and perform random node sampling to simulate variability in arterial density. $H = 3$ is used for hierarchical encoding. The choice of H=3 was empirically validated through ablation studies across $H \in \{1, 2, 3, 4, 5\}$. We found that H=3 provides optimal performance while H>3 shows diminishing returns due to over-coarsening of vascular structure. All experiments are implemented in PyTorch [19] and PyTorch Geometric [7], and executed on an NVIDIA RTX 4090 GPU.

3.2 Results

Regression and Classification Performance. Experimental results demonstrate that all models perform reasonably well on both regression and classification tasks, underscoring the value of our curated dataset. Notably, the HGT, which builds upon the GPS[20], consistently outperforms other baselines in both regression and classification, as shown in Table 1 and Table 2. This consistent improvement highlights the critical role of our hierarchical structural encoding in capturing complex patterns in the BAN.

Attention Visualization. Attention visualization in Fig.3 demonstrates that age can be effectively represented through global BAN structure. Notably, distal arterial branches exhibit higher attention weights, aligning with prior studies [23,4].

Ablation Studies. Table 3 demonstrates our model's effectiveness in FR risk level prediction. Removing hierarchical encoding reduces performance, indicating its importance for capturing multi-scale features. Without the virtual node, information flow is disrupted. Removing BAN features (length, tortuosity, radius) causes severe degradation, confirming these attributes' critical role. Using only bifurcation and endpoint nodes (post topology-preserving reduction) similarly impairs performance. The last two columns show how BAN features contribute significantly to the predictive task and demonstrate the unique value of our comprehensive multimodal dataset.

Table 1: Regression performance across tasks (MAE↓/R^2↑)

Model	Age	HDL	LDL	TC	SBP	DBP	FRS
GCN	4.81±0.38	4.48±0.40	4.74±0.45	4.47±0.50	5.03±0.81	4.49±0.51	4.36±0.27
	0.43±0.08	0.51±0.06	0.46±0.10	0.51±0.09	0.40±0.18	0.51±0.08	0.53±0.05
GAT	2.42±0.28	2.47±0.23	2.37±0.21	1.30±0.05	2.46±0.21	2.42±0.23	2.46±0.20
	0.84±0.03	0.83±0.03	0.84±0.03	0.95±0.00	0.83±0.03	0.84±0.03	0.84±0.03
GT	4.29±0.42	3.95±0.37	4.22±0.48	3.96±0.52	4.57±0.79	3.98±0.49	3.84±0.31
	0.65±0.07	0.72±0.05	0.67±0.11	0.70±0.10	0.62±0.17	0.73±0.09	0.74±0.06
GPS	3.51±0.31	3.22±0.23	3.44±0.37	3.10±0.22	3.68±0.49	3.36±0.32	3.52±0.38
	0.72±0.05	0.78±0.04	0.73±0.05	0.84±0.03	0.79±0.06	0.81±0.05	0.82±0.04
HGT	**2.12**±0.21	**2.15**±0.18	**2.05**±0.15	**1.10**±0.04	**2.27**±0.18	**2.12**±0.20	**2.05**±0.18
(ours)	**0.88**±0.02	**0.87**±0.03	**0.89**±0.03	**0.97**±0.00	**0.86**±0.03	**0.88**±0.02	**0.89**±0.02

4 Discussion & Conclusion

In conclusion, this work advances graph learning in the cardiovascular field by introducing a novel HGT with edge-aware structural encoding that effectively captures the complex structural information within BAN. We have developed the first comprehensive multimodal full-brain BAN dataset enabling robust evaluation across multiple prediction tasks. Our approach demonstrates state-of-the-art performance throughout all predictive tasks, revealing strong associations between vascular structure and cardiovascular risk that conventional methods fail to capture. These innovations collectively enhance our understanding of vascular-systemic relationships while establishing a foundation for more precise, personalized cardiovascular risk stratification that could significantly impact preventive cardiovascular health. While promising, our method may not fully capture the extensive anatomical variability inherent to diseased populations (e.g., aneurysm), potentially limiting robustness in cases with atypical BANs.

Clinical Limitations and Future Directions. While our method demonstrates strong predictive performance, several limitations must be addressed for clinical translation: (1) lack of prospective clinical validation in real workflows, (2) absence of clinician evaluation of risk predictions, and (3) need for integration with existing clinical decision support systems. Future work should focus on prospective clinical trials to validate the clinical utility of BAN-based risk stratification.

Acknowledgments. This study was funded by the National Institutes of Health (grant numbers 5R01NS127317-03 and 5R01NS125635-03).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Table 2: Classification performance across tasks (Micro-F1/Macro-F1/AUC)

Model	Smoking	Diabetes	HT	FR Level
GCN	72.40±4.38	74.38±4.69	74.38±4.14	74.38±4.14
	71.16±5.08	73.19±5.34	73.43±4.65	73.43±4.65
	81.79±2.12	82.51±2.83	82.36±1.53	82.36±1.53
GAT	72.39±3.72	71.37±6.63	71.14±1.90	69.63±3.46
	71.12±4.30	69.41±7.44	69.64±1.54	67.76±2.98
	80.59±3.61	80.06±3.74	79.09±1.39	79.93±3.25
GT	**77.48**±4.41	**79.42**±4.73	**79.45**±4.22	**79.47**±4.19
	76.21±5.12	**78.24**±5.29	**78.51**±4.71	**78.49**±4.68
	82.85±2.18	83.58±2.79	83.41±1.60	83.40±1.55
GPS	74.71±8.51	74.68±10.36	70.68±8.45	74.68±8.31
	72.88±8.40	74.53±10.39	69.90±12.19	72.12±8.69
	77.42±4.12	80.75±10.13	78.75±5.32	79.75±10.92
HGT (ours)	75.64±1.70	76.84±1.45	76.53±1.39	76.23±1.23
	74.82±1.61	75.35±1.42	75.24±1.36	74.92±1.26
	83.23±1.41	**84.12**±1.31	**83.82**±1.23	**83.45**±1.12

Table 3: Ablation studies on FR level prediction (Micro-F1/Macro-F1/AUC)

Metrics	Full Model	w/o Hier. Enc.	w/o Virtual Node	w/o Art. Feat.	Endpts + Bifurc.
Micro-F1	**76.23**±1.23	74.68±8.31	75.53±1.20	52.45±3.41	50.67±1.51
Macro-F1	**74.92**±1.26	72.12±8.69	73.32±1.42	37.89±4.60	36.12±1.72
AUC	**83.45**±1.12	79.75±10.92	81.15±1.78	60.23±5.50	58.46±1.63

References

1. Chen, J., Saad, Y., Zhang, Z.: Graph coarsening: From scientific computing to machine learning. SeMA **79**(1), 187–223 (2022). https://doi.org/10.1007/s40324-021-00282-x
2. Chen, L., Mossa-Basha, M., Balu, N., Canton, G., Sun, J., Pimentel, K., Hatsukami, T.S., Hwang, J.N., Yuan, C.: Development of a quantitative intracranial vascular features extraction tool on 3d mra using semiautomated open-curve active contour vessel tracing. Magnetic Resonance in Medicine **79**(6), 3229–3238 (2018). https://doi.org/10.1002/mrm.26961
3. Chen, L., Mossa-Basha, M., Sun, J., Hippe, D.S., Balu, N., Yuan, Q., Pimentel, K., Hatsukami, T.S., Hwang, J.N., Yuan, C.: Quantification of morphometry and intensity features of intracranial arteries from 3d tof mra using the intracranial artery feature extraction (icafe): A reproducibility study. Magnetic Resonance Imaging **57**, 293–302 (2019). https://doi.org/10.1016/j.mri.2018.12.007
4. Chen, L., Sun, J., Hippe, D.S., Balu, N., Yuan, Q., Yuan, I., Zhao, X., Li, R., He, L., Hatsukami, T.S., Hwang, J.N., Yuan, C.: Quantitative assessment of the intracranial vasculature in an older adult population using icafe. Neurobiology of Aging **79**, 59–65 (2019). https://doi.org/10.1016/j.neurobiolaging.2019.02.027

5. Dwivedi, V.P., Bresson, X.: A generalization of transformer networks to graphs. arXiv preprint arXiv:2012.09699 (2021). https://doi.org/10.48550/arXiv.2012.09699
6. D'Agostino, R.B., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M., Kannel, W.B.: General cardiovascular risk profile for use in primary care. Circulation **117**(6), 743–753 (2008). https://doi.org/10.1161/CIRCULATIONAHA.107.699579
7. Fey, M., Lenssen, J.E.: Fast graph representation learning with pytorch geometric. arXiv preprint arXiv:1903.02428 (2019). https://doi.org/10.48550/arXiv.1903.02428
8. Guo, Y., Canton, G., Chen, L., Sun, J., Geleri, D.B., Balu, N., Xu, D., Mossa-Basha, M., Hatsukami, T.S., Yuan, C.: Multi-planar, multi-contrast and multi-time point analysis tool (mocha) for intracranial vessel wall characterization. Journal of Magnetic Resonance Imaging **56**(3), 944–955 (2022). https://doi.org/10.1002/jmri.28087
9. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference (SciPy2008). pp. 11–15 (2008)
10. Huang, F., Lian, J., Ng, K.S., Shih, K., Vardhanabhuti, V.: Predicting ct-based coronary artery disease using vascular biomarkers derived from fundus photographs with a graph convolutional neural network. Diagnostics **12**(6), 1390 (2022). https://doi.org/10.3390/diagnostics12061390
11. Hwang, E., Thost, V., Dasgupta, S.S., Ma, T.: An analysis of virtual nodes in graph neural networks for link prediction (extended abstract). In: The First Learning on Graphs Conference (2022), https://openreview.net/forum?id=dI6KBKNRp7
12. Li, H., Wang, M., Ma, T., Liu, S., Zhang, Z., Chen, P.Y.: What improves the generalization of graph transformers? a theoretical dive into the self-attention and positional encoding. arXiv preprint arXiv:2406.01977 (2024). https://doi.org/10.48550/arXiv.2406.01977
13. Liu, W., Huang, X., Liu, X., Ortega, D., Chen, L., Chen, Z., Sun, J., Wang, L., Hatsukami, T.S., Yuan, C., Li, H., Yang, J.: Uncontrolled hypertension associates with subclinical cerebrovascular health globally: A multimodal imaging study. European Radiology **31**(4), 2233–2241 (2021). https://doi.org/10.1007/s00330-020-07218-5
14. Luo, Y., Li, H., Shi, L., Wu, X.M.: Enhancing graph transformers with hierarchical distance structural encoding. arXiv preprint arXiv:2308.11129 (2024). https://doi.org/10.48550/arXiv.2308.11129
15. Luo, Y., Shi, L., Wu, X.M.: Classic gnns are strong baselines: Reassessing gnns for node classification (2024), https://arxiv.org/abs/2406.08993
16. Microsoft: Neural network intelligence (2021), version 2.10
17. Mishra, S., Wang, Y.X., Wei, C.C., Chen, D.Z., Hu, X.S.: VTG-Net: A cnn based vessel topology graph network for retinal artery/vein classification. Frontiers in Medicine **8**, 750396 (2021). https://doi.org/10.3389/fmed.2021.750396
18. Paetzold, J.C., McGinnis, J., Shit, S., Ezhov, I., Büschl, P., Prabhakar, C., Todorov, M.I., Sekuboyina, A., Kaissis, G., Ertürk, A., Günnemann, S., Menze, B.H.: Whole brain vessel graphs: A dataset and benchmark for graph learning and neuroscience. arXiv preprint arXiv:2108.13233 (2022). https://doi.org/10.48550/arXiv.2108.13233
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z.,

Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library. arXiv preprint arXiv:1912.01703 (2019). https://doi.org/10.48550/arXiv.1912.01703
20. Rampá, L., Galkin, M., Dwivedi, V.P., Luu, A.T., Wolf, G., Beaini, D.: Recipe for a general, powerful, scalable graph transformer. arXiv preprint (2021)
21. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2018). https://doi.org/10.48550/arXiv.1710.10903
22. Zhang, K., Akcicek, H., Shi, G., Canton, G., Liu, J., Guo, Y., Wang, X., Chen, L., Pimentel, K.D., Akcicek, E.Y., Tang, X., Jin, Y., Li, X., Balu, N., Hatsukami, T.S., Mossa-Basha, M., Chen, Z., Yuan, C.: Estimating flow direction of circle of willis using dynamic arterial spin-labeling mr angiography. AJNR American Journal of Neuroradiology **45**(10), 1419–1426 (2024). https://doi.org/10.3174/ajnr.A8355
23. Zhang, K., Chen, Z., Chen, L., Canton, G., Geleri, D.B., Chu, B., Guo, Y., Hippe, D.S., Pimentel, K.D., Balu, N., Hatsukami, T.S., Yuan, C.: Alterations in cerebral distal vascular features and effect on cognition in a high cardiovascular risk population: A prospective longitudinal study. Magnetic Resonance Imaging **98**, 36–43 (2023). https://doi.org/10.1016/j.mri.2022.12.011
24. Zhang, K., Gould, A., Chen, L., Chen, Z., Canton, G., Balu, N., Hatsukami, T.S., Yuan, C.: Vascular map assessment—the dual use of an analysis tool to evaluate imaging technique reproducibility and AI technique's performance. In: Proceedings of the 2022 ISMRM & ISMRT Annual Meeting & Exhibition. p. Abstract #4934 (2022)
25. Zhang, Z., Liu, Q., Hu, Q., Lee, C.K.: Hierarchical graph transformer with adaptive node sampling. arXiv preprint arXiv:2210.03930 (2022). https://doi.org/10.48550/arXiv.2210.03930