

Domain-Adaptive Diagnosis of Lewy Body Disease with Transferability Aware Transformer

Xiaowei Yu¹, Jing Zhang¹, Tong Chen¹, Yan Zhuang¹, Minheng Chen¹, Chao Cao¹, Yanjun Lyu¹, Lu Zhang², Li Su^{3,4}, Tianming Liu⁵, and Dajiang Zhu¹✉

¹ Department of Computer Science and Engineering, University of Texas at Arlington, USA

{dajiang.zhu@uta.edu}

² Department of Computer Science, Indiana University Indianapolis, USA

³ Sheffield Institute for Translational Neuroscience, University of Sheffield, UK

⁴ Department of Psychiatry, University of Cambridge, UK

⁵ School of Computing, University of Georgia, USA

Abstract. Lewy Body Disease (LBD) is a common yet understudied form of dementia that imposes a significant burden on public health. It shares clinical similarities with Alzheimer’s disease (AD), as both progress through stages of normal cognition, mild cognitive impairment, and dementia. A major obstacle in LBD diagnosis is data scarcity, which limits the effectiveness of deep learning. In contrast, AD datasets are more abundant, offering potential for knowledge transfer. However, LBD and AD data are typically collected from different sites using different machines and protocols, resulting in a distinct domain shift. To effectively leverage AD data while mitigating domain shift, we propose a Transferability Aware Transformer (TAT) that adapts knowledge from AD to enhance LBD diagnosis. Our method utilizes structural connectivity (SC) derived from structural MRI as training data. Built on the attention mechanism, TAT adaptively assigns greater weights to disease-transferable features while suppressing domain-specific ones, thereby reducing domain shift and improving diagnostic accuracy with limited LBD data. The experimental results demonstrate the effectiveness of TAT. To the best of our knowledge, this is the first study to explore domain adaptation from AD to LBD under conditions of data scarcity and domain shift, providing a promising framework for domain-adaptive diagnosis of rare diseases.

Keywords: Lewy Body Disease · Domain Adaptation · Transferability Aware Transformer.

1 Introduction

Lewy Body Disease (LBD) is a progressive neurodegenerative disorder and one of the most common causes of dementia [1], yet it remains underexplored compared to Alzheimer’s disease (AD). LBD shares clinical similarities with AD, as both diseases exhibit cognitive decline progressing through cognitive normal

(CN), mild cognitive impairment (MCI), and dementia stages. For both LBD and AD, once established, they cannot be cured or prevented [2][4]. Therefore, improving diagnostic accuracy is crucial for early intervention to slow disease progression [3][5][6]. Despite this importance, accurate diagnosis of LBD remains a challenge due to the scarcity of LBD data, limiting the effectiveness of deep learning models.

In contrast, Alzheimer’s disease has been extensively studied, with large-scale datasets publicly available [7][8][10][12][13]. This discrepancy presents an opportunity for knowledge transfer [9], where insights gained from AD data can enhance LBD diagnosis. However, a key obstacle in this process is the domain shift, where LBD and AD datasets are usually collected from different institutions, using varying imaging machines and acquisition protocols, leading to distinct differences in data distribution. Conventional deep learning models struggle with such discrepancies, limiting their generalizability across diseases.

To leverage knowledge from a larger AD dataset for effective model training while mitigating the domain shift, we propose the Transferability Aware Transformer (TAT), a novel vision transformer (ViT)-based approach designed to adapt knowledge from AD to improve LBD diagnosis. TAT utilizes structural connectivity (SC) derived from structural MRI as input and selectively emphasizes transferable features while suppressing domain-specific ones. At the core of TAT is the Transferability Aware Self-Attention (TAS) mechanism, specifically designed for ViT to amplify disease-transferable features between LBD and AD while suppressing less transferable ones. TAS employs a local discriminator to assess the transferability of patches by distinguishing whether they originate from LBD or AD datasets, generating a transferability matrix. This adaptively learned matrix is then integrated into the self-attention mechanism, enabling the model to focus effectively on transferable features. Through this process, TAS assigns greater weights to more transferable attention operations while reducing the influence of less transferable features. To further mitigate domain shift, we introduce a global discriminator, which takes the classification token as input to distinguish samples from LBD and AD datasets. By aligning cross-domain representations, TAT leverages the larger AD dataset to enhance LBD diagnosis. Experimental results demonstrate that TAT outperforms state-of-the-art domain adaptation methods, providing a promising framework for LBD diagnosis.

2 Method

We model the problem within the domain adaptation framework [11][14]. Let $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ represent samples in the labeled AD dataset, where x_i^s denotes the structural connectivity (SC), y_i^s denotes the corresponding labels, and n_s is the number of samples. Similarly, let $D_t = \{(x_j^t)\}_{j=1}^{n_t}$ represent samples in the LBD dataset, consisting of n_t SCs but no labels. Since the number of available LBD samples is limited, it satisfies $n_t \ll n_s$. Our proposed TAT aims to learn transferable knowledge to minimize domain discrepancy, thereby achieving the desired prediction performance on LBD dataset.

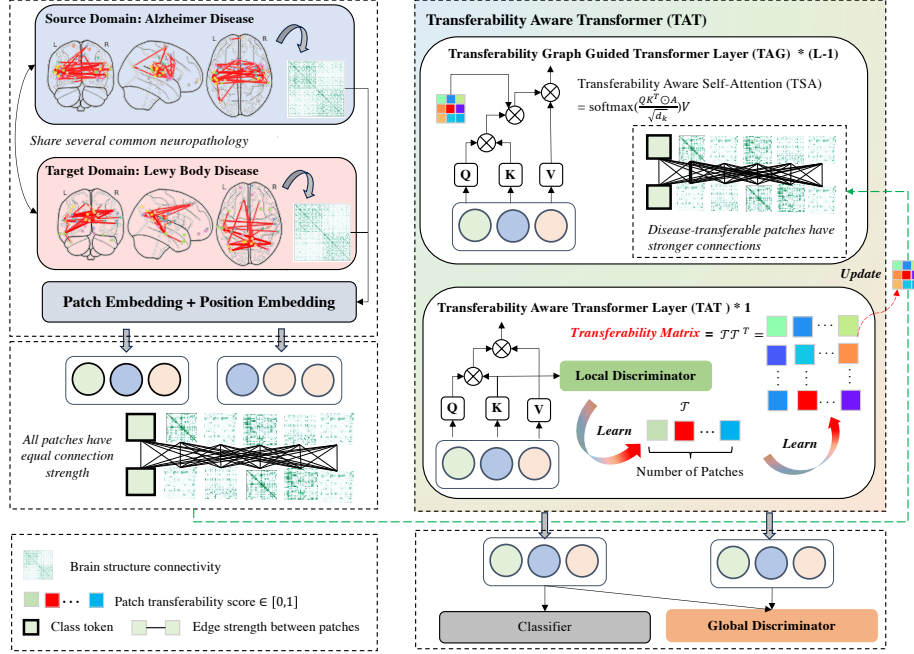


Fig. 1: Illustration of proposed framework. The TAT layer employs a local discriminator to dynamically assess the transferability score of each SC patch and generate a corresponding transferability matrix. This matrix is then integrated into other Transformer layers to adjust the self-attention strength and thus enhance cross-domain knowledge sharing. The global discriminator utilizes the classification token to align the global features.

2.1 Overview of TAT

Figure 1 illustrates the overall framework of our proposed TAT. TAT is built upon a Vision Transformer (ViT) and comprises a global discriminator that evaluates the entire SCs using class tokens, a local discriminator that assesses the transferability of individual patches, and a classification head. The ViT backbone is enhanced with the transferability-aware attention, which includes both the Transferability Graph-Guided Transformer Layer and the Transferability-Aware Transformer Layer.

Following the patch division in ViT [15][16], each SC is divided into non-overlapping patches, which are linearly projected into the latent space and concatenated with positional information. A class token is prepended to the patch tokens, and all tokens are subsequently processed by the backbone. From a graph perspective, we consider the patches as nodes and the attention between patches as edges. This allows us to manipulate the strength of self-attention by adaptively learning a transferability-aware graph (TAG).

2.2 Transferability Aware Transformer Layer

The patch tokens correspond to partial of the SC and capture features as fine-grained brain representations. In this work, we define the transferability score on a patch to assess its transferability. A patch with a higher transferability score is more likely to correspond to the highly transferable features across the diseases.

To obtain the transferability score of patch tokens, we adopt a patch-level local discriminator D_l to evaluate the local features with a loss function:

$$L_{pat}(\tilde{z}, y^d) = -\frac{1}{nP} \sum_{\tilde{z}_i \in D} \sum_{p=1}^P L_{CE}(D_l(\tilde{z}_{ip}), y_{ip}^d) \quad (1)$$

where P is the number of patches, $D = \{D_s\} \cup \{D_t\}$, D_l is the local discriminator, $n = \sum_{s=1}^S n_s + \sum_{t=1}^T n_t$, is the total number of SC in D , y_{ip}^d denotes the macro label of the p th token of the i th SC, i.e., $y_{ip}^d = 1$ means labeled SCs in AD dataset, $y_{ip}^d = 0$ the unlabeled LBD dataset. The distributions of features are aligned when the local discriminator finds it difficult to differentiate the macro labels of the patch tokens. $D_l(\tilde{z}_{ip})$ gives the probability of the patch belonging to the labeled samples in the AD dataset. During the adaptation process, D_l tries to discriminate the patches correctly, assigning 1 to patches from the labeled data and 0 to those from the unlabeled data.

Empirically, patches that cannot be easily distinguished by the local discriminator (e.g., D_l is around 0.5) are more likely to correspond to highly transferable features across different diseases, as they capture underlying patterns that remain consistent despite variations in the data distribution. We use the phrase ‘‘transferability’’ to capture this property. We define the transferability score of a patch as:

$$c(\tilde{z}_{ip}) = H(D_l(\tilde{z}_{ip})) \in [0, 1] \quad (2)$$

where $H(\cdot)$ is the standard entropy function. If the output of the local discriminator D_l is around 0.5, then the transferability score is close to 1, indicating that the features in the patch are highly transferable. This enables us to quantify the transferability of patches. Assessing the transferability of patches allows a finer-grained view of the SC, separating an SC into highly transferable and less transferable patches. Features from highly transferable patches will be amplified while features from less transferable patches will be suppressed.

Let $\mathcal{C}_i = \{c_{i1}, \dots, c_{ip}\}$ be the transferability scores of patches of SC i . The transferability matrix, i.e., the adjacency matrix of the learned transferability-aware graph (TAG), can be formulated as:

$$A = \frac{1}{B\mathcal{H}} \sum_{h=1}^{\mathcal{H}} \sum_{i=1}^B [\mathcal{C}_i^T \mathcal{C}_i]_{\times} \quad (3)$$

where B is the batch size, $\mathcal{H}$ is the number of heads, $[\cdot]_{\times}$ means no gradients back-propagation for the transferability matrix. The adjacency matrix of learned TAG A controls the connection strength of attention between patches.

Since the Transferability Aware Transformer Layer is the final layer in the model, the class token needs to consider the transferability scores and emphasize these high-transferability features from patch tokens. Consequently, the vanilla Self-Attention (SA) in this layer can be upgraded to Transferability Aware Self-Attention (TAS) by incorporating the transferability scores

$$TAS(q_{cls}, K, V) = softmax(\frac{q_{cls}K^T}{\sqrt{d}}) \odot [1; C(K_{patch})] V \quad (4)$$

where q_{cls} is the query of the class token, K represents the key of all tokens, including the class token and patch tokens, K_{patch} is the key of the patch tokens, $C(K_{patch})$ denotes the transferability scores of the patch tokens, $\odot$ is the dot product, and $[;]$ is the concatenation operation. The new self-attention encourages the class token to take more information from highly transferable patches while suppressing information from patches with low transferability scores. Accordingly, the multi-head self-attention is therefore renewed as:

$$MTAS(q_{cls}, K, V) = Concat(head_1, \dots, head_h)W^O \quad (5)$$

where $head_i = TAS(q_{cls}W_i^{q_{cls}}, KW_i^K, VW_i^V)$, as formulated in Eq. 4. Taking them together, the operations in the transferability aware transformer layer can be formulated as:

$$\begin{aligned} \hat{z}^l &= MTAS(LN(\hat{z}^{l-1})) + \hat{z}^{l-1} \\ z^l &= MLP(LN(\hat{z}^l)) + \hat{z}^l \end{aligned} \quad (6)$$

where z^{l-1} are output from the previous layer. In this way, the Transferability Aware Transformer Layer produces transferability scores for each patch token while also emphasizing fine-grained features that are highly transferable and discriminative for classification. Here $l = L$, L is the total number of transformer layers in ViT architecture.

2.3 Transferability Graph Guided Transformer Layer

The adaptively learned TAG can be conveniently integrated to guide self-attention by applying its adjacency matrix via dot product with the attention scores, as depicted in the Transferability Graph Guided Transformer Layer in Fig. 1. The TAG guided self-attention focuses on the patches with more transferable features, thus steering the model to capture transferable knowledge. Formally, we can formulate the TAG guided self-attention as:

$$TAG-SA(Q, K, V, A) = softmax(\frac{QK^T \odot A}{\sqrt{d_k}})V \quad (7)$$

where queries, keys, and values of all patches are packed into matrices Q , K , and V , respectively, A is the adjacency matrix of the learned TAG. Accordingly, the TAG guided multi-head attention is then formulated as:

$$MTAG-SA(Q, K, V, A) = Concat(head_1, \dots, head_h)W^O \quad (8)$$

where $head_i = TAG-SA(QW_i^Q, KW_i^K, VW_i^V, A)$. The learnable parameter matrices W_i^Q , W_i^K , W_i^V and W^O are the projections. The new knowledge learned from the unseen data is incorporated into the model by integrating the learned matrix with the attention scores.

2.4 Overall Objective Function

Since our proposed TAT has a classifier head, a local discriminator, and a global discriminator, there are three terms in the overall objective function. The classification loss is formulated as:

$$L_{clc}(x^s, y^s) = \frac{1}{n_s} \sum_{x_i \in D_s} L_{CE}(G_c(G_f(x_i^s)), y_i^s) \quad (9)$$

where G_c is the classifier head, and G_f is the feature extractor of ViT. The global discriminator takes the class token of images and tries to discriminate the class token, i.e., the representation of the entire image, to the labeled support data and unlabeled testing data:

$$L_{dis}(x, y^d) = -\frac{1}{n} \sum_{x_i \in D} L_{ce}(D_g(G_f(x_i), y_i^d)) \quad (10)$$

where D_g is the global discriminator, and y_i^d is the macro label (i.e., $y_i^d = 1$ means labeled support data, $y_i^d = 0$ is the unlabeled testing data). Take classification loss (Eq. 9), global discrimination loss (Eq. 10), and local discrimination loss (Eq. 1) together, the overall objective function is therefore formulated as:

$$L = L_{clc}(x^s, y^s) + \alpha L_{dis}(x, y^d) + \beta L_{pat}(x, y^d) \quad (11)$$

where α and β are trainable hyperparameters.

2.5 Open-set Adaptation

The label categories in the source dataset include CN, MCI, and AD. However, the label categories in the target dataset consist of CN, MCI, and LBD. This creates an open-set challenge, as the LBD dataset contains a disease category not present in the source dataset.

To address this challenge, we use CN and MCI subjects from the source dataset and include CN, MCI, and LBD subjects from the target dataset during training. Since the labels of subjects from the source dataset are provided during training, the classifier is trained as a binary classifier to distinguish between CN and MCI. To identify the additional LBD disease, we propose a threshold-based entropy approach. For a given subject, the classifier outputs a probability distribution $P = [p_{CN}, p_{MCI}]$, where p_{CN} and p_{MCI} represent the predicted probabilities for CN and MCI, respectively. We compute the entropy of the prediction as:

$$H(P) = -p_{CN} \log p_{CN} - p_{MCI} \log p_{MCI} \quad (12)$$

Table 1: Comparison between TAT and other models on the LBD dataset. Accuracy and standard deviation are reported based on three runs. ViT, TVT, and SSRT only support closed-set adaptation.

Method	CN	MCI	LBD
ViT	4.3 ± 3.5	88.3 ± 13.5	N/A
TVT	46.4 ± 5.4	16.7 ± 13.5	N/A
SSRT	47.9 ± 3.4	11.1 ± 7.9	N/A
TAT	66.7 ± 7.4	88.9 ± 7.9	14.5 ± 4.3

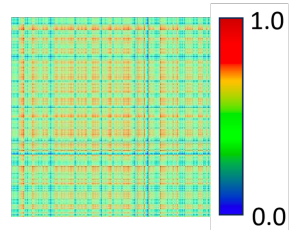


Fig. 2: The adjacency matrix of the learned TAG.

where $H(P)$ quantifies the uncertainty of the prediction. If the entropy exceeds a predefined threshold τ , the sample is considered to belong to the unknown LBD category:

$$\text{Predicted Label} = \begin{cases} \text{LBD}, & \text{if } H(P) > \tau \\ \arg \max(P), & \text{otherwise} \end{cases} \quad (13)$$

3 Experiments

3.1 Datasets and Training Details

We use the structural connectivity (SC) constructed from MRI images, and the pre-processing follows the procedures established in existing works [17]. For the AD dataset, we use the public Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset, which is the most widely used Alzheimer’s disease dataset [18]. Through SC construction, 282 CN and 149 MCI subjects are available. The LBD dataset was collected using different machines and protocols, resulting in a distinct domain shift from the ADNI dataset. There are 23 CN, 6 MCI, and 77 LBD subjects available, which is significantly fewer than those in ADNI.

Since the datasets are relatively small in scale, we build our TAT model on ViT-Small (ViT-S) and initialize it with pre-trained weights [19][20]. We initialize the learning rate at 0 and linearly warm it up to 0.06 over the first 500 training steps, followed by a cosine decay scheduling strategy [21][22]. The hyperparameters α and β are set to [1.0, 0.01]. The entropy threshold is set to 0.8. We use a mini-batch Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.9. The batch size is set to 16. During training, the model is trained on both the ADNI and LBD datasets. The ADNI samples are provided with labels, whereas the LBD samples are unlabeled. After convergence, we evaluate the prediction results on the LBD dataset against the true labels. This approach enables TAT to adapt to unseen distributions during training.

Table 2: Ablation study of hyperparameters of α , β , and τ . Accuracy and standard deviation are based on three runs.

(α, β, τ)	CN	MCI	LBD
(1.0,0.1,0.8)	57.9 ± 7.4	88.3 ± 13.6	21.3 ± 5.6
(0.1,0.01,0.8)	60.7 ± 14.1	88.3 ± 13.6	14.5 ± 4.3
(1.0,0.01,0.7)	47.8 ± 3.6	77.7 ± 7.9	54.1 ± 8.1

Table 3: Ablation study of global discriminator (GD) and local discriminator (LD). Accuracy and standard deviation are based on three runs.

Method	CN	MCI	LBD
TAT	66.7 ± 7.4	88.9 ± 7.9	14.5 ± 4.3
w/o GD	50.8 ± 5.4	66.7 ± 13.6	12.9 ± 2.8
w/o LD	49.3 ± 14.7	88.3 ± 13.6	14.2 ± 4.8

3.2 Adaptation Results

The adaptation results are reported in Table 1. We compare the proposed TAT with other Transformer-based models. ViT serves as the baseline, while TVT [23] and SSRT [24] represent the latest advancements in domain adaptation. However, ViT, TVT, and SSRT only support closed-set adaptation, where the label categories are required to be the same between training and inference. Our proposed TAT not only outperforms these competitors in the MCI and CN categories, but is also able to distinguish the LBD subjects.

We also visualize the learned transferability-aware graph (TAG) to better understand the transferability of the SC patches. The TAG is visualized in the form of the symmetric adjacency matrix, which is shown in Fig. 2. The transferability scores for each SC patch are represented by the intensity of red colors (increasing) and blue colors (decreasing). We observe that different SC patches have different transferability scores. Certain SC patches have higher transferability scores and therefore contain transferable features.

3.3 Ablation Study

To evaluate the influence of the hyperparameters and key modules on model performance, we conduct an ablation study. The results are reported in Table 2 and 3. We observe in Table 2 that the model is robust to different combinations of hyperparameters given the limited data samples. A lower entropy threshold leads to higher accuracy in the LBD category, but decreases classification accuracy in others. The results in Table 3 show that both the global discriminator and local discriminator contribute to the final performance.

4 Conclusion

In this work, we propose a novel domain adaptation framework for diagnosing LBD under scarce data and domain shift conditions. TAT addresses two key challenges in LBD analysis: first, it leverages knowledge from the clinically similar AD dataset to enhance model effectiveness; second, it effectively mitigates the distribution shift between datasets collected from different institutions using various machines and protocols. Experimental results demonstrate that TAT outperforms existing domain adaptation methods, achieving robust prediction results for LBD diagnosis.

Acknowledgments. This work was supported by the National Institutes of Health (R01AG075582 and RF1NS128534) and funded by the Sir Jules Thorn Charitable Trust (05/JTA), Wellcome Trust Intermediate Clinical Fellowship (WT088441MA), Alzheimer’s Research UK Senior Research Fellowship (ARUK-SRF2017B-1), The Lewy body Society (LBS/002/2019) and NIHR Biomedical Research Centres in Sheffield, Cambridge and Newcastle. We also thank Andrew M. Blamire, John-Paul Taylor, Li Su and John T. O’Brien for sharing the data.

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

References

1. Walker, Z., Possin, K. L., Boeve, B. F. and Aarsland, D.: Lewy body dementias. *The Lancet* **386**(10004), 1683-1697 (2015)
2. Lyu, Y., Yu, X., Zhu, D. and Zhang, L.: Classification of alzheimer’s disease via vision transformer. In Proceedings of the 15th international conference on Pervasive technologies related to assistive environments, pp. 463-468. (2022)
3. Lyu, Y., Zhang, L., Yu, X., Cao, C., Liu, T. and Zhu, D.: Mild Cognitive Impairment Classification Using A Novel Finer-Scale Brain Connectome. In IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1-5. (2024)
4. Capouch, S. D., Farlow, M. R. and Brosch, J. R.: A review of dementia with Lewy bodies’ impact, diagnostic criteria and treatment. *Neurology and Therapy* **7**, 249-263 (2018)
5. Zhang, J., Lyu, Y., Yu, X., Zhang, L., Cao, C., Chen, T., Chen, M., Zhuang, Y., Liu, T. and Zhu, D.: Classification of Mild Cognitive Impairment Based on Dynamic Functional Connectivity Using Spatio-Temporal Transformer. arXiv preprint arXiv:2501.16409 (2025)
6. Zhang, J., Yu, X., Lyu, Y., Zhang, L., Chen, T., Cao, C., Zhuang, Y., Chen, M., Liu, T. and Zhu, D.: Brain-Adapter: Enhancing Neurological Disorder Analysis with Adapter-Tuning Multimodal Large Language Models. arXiv preprint arXiv:2501.16282 (2025)
7. Yu, X., Scheel, N., Zhang, L. Zhu, D.C, Zhang, R. and Zhu, D.: Free water in T2 FLAIR white matter hyperintensity lesions. *Alzheimer’s & Dementia*, e057398 (2021)
8. Zhang, J., Yu, X., Chen, T., Cao, C., Chen, M., Zhuang, Y., Lyu, Y., Zhang, L., Su, L., Liu, T. and Zhu, D.: BrainNet-MoE: Brain-Inspired Mixture-of-Experts Learning for Neurological Disease Identification. arXiv preprint arXiv:2503.07640 (2025)
9. Fang, L., Yu, X., Cai, J., Chen, Y., Wu, S., Liu, Z., Yang, Z., Lu, X., Gong, X., Liu, Y., Ma, T., Ruan, W., Abbasi, A., Zhang, J., Wang, T., Latif, E., Liu, W., Zhang, W., Kolouri, S., Zhai, X., Zhu, D., Zhong, W., Liu, T. and Ma, P.: Knowledge Distillation and Dataset Distillation of Large Language Models: Emerging Trends, Challenges, and Future Directions. arXiv preprint arXiv:2504.14772 (2025)
10. Lyu, Y., Wu, Z., Zhang, L., Zhang, J., Li, Y., Ruan, W., Liu, Z., Yu, X., Cao, C., Chen, T., Chen, M., Zhuang, Y., Li, X., Liu, R., Huang, C., Li, W., Liu, T. and Zhu, D.: Gp-gpt: Large language model for gene-phenotype mapping. arXiv preprint arXiv:2409.09825 (2024)
11. Yu, X., Huang, Z. and Zhang, Z.: Feature fusion transferability aware transformer for unsupervised domain adaptation. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV), pp. 6752-6761. (2025)

12. Lyu, Y., Yu, X., Zhang, L. and Zhu, D.: Classification of mild cognitive impairment by fusing neuroimaging and gene expression data. In Proceedings of the 14th Pervasive Technologies Related to Assistive Environments Conference, pp. 26-32. (2021)
13. Chen, M., Yu, X., Zhang, J., Chen, T., Cao, C., Zhuang, Y., Lyu, Y., Zhang, L., Liu, T. and Zhu, D.: Core-Periphery Principle Guided State Space Model for Functional Connectome Classification. arXiv preprint arXiv:2503.14655 (2025)
14. Yu, X., Zhu, D. and Liu, T.: Robust core-periphery constrained transformer for domain adaptation. arXiv preprint arXiv:2308.13515 (2023)
15. Yu, X., Zhang, L., Wu, Z. and Zhu, D.: Core-Periphery Multi-Modality Feature Alignment for Zero-Shot Medical Image Analysis. IEEE Transactions on Medical Imaging (2024)
16. Yu, X., Wu, Z., Zhang, L., Zhang, J., Lyu, Y. and Zhu, D.: Cp-clip: Core-periphery feature alignment clip for zero-shot medical image analysis. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 88-97. (2024)
17. Zhang, L., Yu, X., Lyu, Y., Liu, T. and Zhu, D.: Representative Functional Connectivity Learning for Multiple Clinical groups in Alzheimer's Disease. In IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1-5. (2023)
18. Yu, X., Zhang, L., Lyu, Y., Liu, T. and Zhu, D.: Supervised deep tree in alzheimer's disease. In IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1-5. (2023)
19. Huang, Z., Yu, X., Zhu, D. and Hughes, M.C.: Interlude: Interactions between labeled and unlabeled data to enhance semi-supervised learning. arXiv preprint arXiv:2403.10658 (2024)
20. Shu, Z., Yu, X., Wu, Z., Jia, W., Shi, Y., Yin, M., Liu, T., Zhu, D. and Niu, W.: Real-time core-periphery guided ViT with smart data layout selection on mobile devices. Advances in Neural Information Processing Systems, pp. 95744-95763. (2024)
21. Yu, X., Huang, Z., Xue, Y., Zhang, L., Wang, L., Liu, T. and Zhu, D.: Noisyenn: Exploring the influence of information entropy change in learning systems. arXiv preprint arXiv:2309.10625 (2023)
22. Yu, X., Zhang, L., Dai, H., Lyu, Y., Zhao, L., Wu, Z., Liu, D., Liu, T. and Zhu, D.: Core-periphery principle guided redesign of self-attention in transformers. arXiv preprint arXiv:2303.15569 (2023)
23. Yang, J., Liu, J., Xu, N. and Huang, J.: Tvt: Transferable vision transformer for unsupervised domain adaptation. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV), pp. 520-530. (2023)
24. Sun, T., Lu, C., Zhang, T. and Ling, H.: Safe self-refinement for transformer-based domain adaptation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 7191-7200. (2022)