Supplementary Material: Cross-domain Dual-stream Feature Disentanglement for Brain Disorder Prediction with Sparsely Labeled PET

1. Model Pseudocode

Algorithm 1 DSDA Framework

Input: source data D_s, labeled target data D_t, unlabeled target data D_u.

Output: feature extractor F, encoder(Graph conv) E, attention network Att, classifier C.

while not converged **do**
  PNIA
  $G_s = (Z_s, A_s)$, $G_t = (Z_t, A_t)$, $G_u = (Z_u, A_u) \leftarrow E\left(F\left(D_s, D_t, D_u\right)\right)$
  $L_{con} \leftarrow g_s, g_u \leftarrow Att\left(Z_s, Z_u\right)$
  DPDFP
  Decoupling and masking non-critical brain regions M and critical brain regions M_s
  $L_{NTA} \leftarrow$ **NTA** $\leftarrow A_s, A_u, Z_s, Z_u, M$
  $L_{pseudo} \leftarrow$ **CHCF** $\leftarrow Z_s, Z_u, M_s$
  $L_s, L_t \leftarrow C\left(Z_s, Z_t\right)$
  Update F, E, Att and C
end while

2. Proportion of Labeled PET

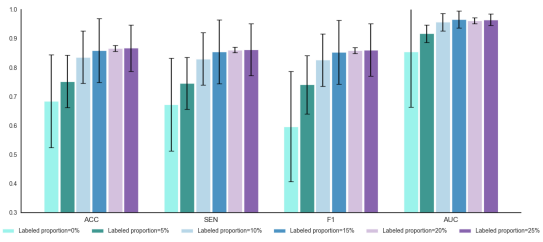
We systematically investigate the impact of varying the proportion of labeled data on model performance under a semi-supervised learning framework, using the ADNI, AIBL, and PPMI datasets. As clearly illustrated in Fig. 1, the model performance exhibits a consistent upward trend as the proportion of labeled data increases. This phenomenon strongly confirms the fundamental role of labeled data as a supervisory signal—each additional labeled sample contributes to a more robust learning foundation, offering richer discriminative cues that effectively guide the model to distill more informative and classification-relevant representations from the vast pool of unlabeled data, thereby steadily enhancing the overall classification accuracy and generalization capability.

However, the magnitude of model performance improvement exhibits a significant decreasing trend. When the label ratio is 0% (i.e., the unsupervised domain adaptation scenario), the performance on all three datasets remains at a relatively low level. As the label ratio gradually increases, the performance is significantly improved; however, when the label ratio rises from 20% to 25%, the performance growth rate slows down remarkably. This behavior can be attributed to two key factors: firstly, at lower annotation levels, newly added labeled samples tend to cover previously unseen critical pathological patterns that are crucial for distinguishing between different disease stages or subtypes, which contribute substantially to improving model performance. As the amount of labeled data increases, the new samples exhibit high redundancy with previously seen patterns, offering limited additional discriminative value. Moreover, the model's learning capacity is inherently bounded by its architectural design, including the number of layers, the number of parameters, and the complexity of its feature extraction mechanisms. Once the labeled data volume reaches a certain threshold, the model approaches its performance ceiling, and further increases in labeled data yield diminishing returns. These observations collectively suggest the existence of a clear cost-performance inflection point in semi-supervised learning. Below this threshold, the investment in labeling more data yields substantial performance gains, making it a cost-effective strategy. Beyond this point, however, the marginal benefit of additional annotations becomes limited, while the cost of data labeling continues to accumulate. At this stage, efforts should shift from increasing label quantity to optimizing model architecture or introducing strategies such as active learning, rather than relying solely on additional annotated samples.

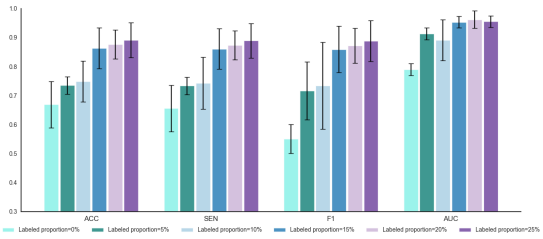
3. Hyperparameter Analysis

We conducted an analysis of the hyperparameters K and α. The hyperparameter K controls the number of selected critical brain regions, while α serves as a trade-off coefficient to balance the supervised and unsupervised loss components. The experimental results are summarized in Tab. 1.

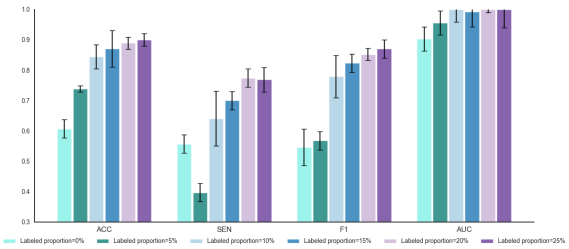
Hyperparameter K. On the two AD datasets (ADNI and AIBL), the model achieved optimal performance when the number of critical brain regions was set to 15. On the PD dataset (PPMI), optimal performance was observed when



(a) ADNI



(b) AIBL



(c) PPMI

Figure 1. Comparison of the diagnostic performance under different proportions of labeled data on three datasets

K = 10. This finding validates the importance of decoupling critical and non-critical brain regions in the proposed framework. When K is too small, discriminative features in critical brain regions may be lost during the feature alignment process, thereby reducing diagnostic accuracy. Conversely, when K is too large, the high-confidence feature fusion process may incorporate excessive noise from non-critical regions, degrading performance. Additionally, the difference in optimal K values between AD and PD is consistent with the distinct neuropathological mechanisms of these diseases. PD pathology (e.g., α-synuclein aggregation) primarily spreads through neural projections in a stepwise manner, initially affecting motor circuits and remaining localized within specific neural pathways. In contrast, AD pathology (e.g., β-amyloid plaques and tau tangles) can rapidly propagate across multiple brain regions via cerebrospinal fluid circulation or neural fiber networks, broadly impacting cognitively relevant cortical areas. As a result, AD tends to involve a greater number of affected brain regions than PD.

Hyperparameter α. Experimental results across the three datasets show that as α increases from 0 to 0.2, performance metrics gradually improve. However, further increasing α beyond this point leads to a decline in performance. Our analysis indicates that when α is too small, the model relies primarily on the supervised loss. Given that labeled PET are significantly fewer compared to labeled MRI, the model becomes overly dependent on the supervision from the MRI modality. This imbalance limits the model's ability to reduce domain discrepancies, resulting in lower prediction accuracy. On the other hand, when α is too large, the model is predominantly constrained by the unsupervised loss. Due to the inherently high uncertainty of unsupervised learning, this introduces instability into the training process. This instability is also reflected in the increased variance of the evaluation metrics when α is large, as shown in Tab. 1.

4. Analysis of Training Consumption

In this section, we compare the training performance of the proposed model with that of FixMatch [5], CAN[7], CDAC[2], CLDA[4], MPL[3], DeCoTa[8], and DCC[1] models, and conduct a detailed analysis of the parameter quantities and training time of each model. The experimental results are illustrated in Figs. 2 and 3.

In Fig. 2, we present the statistics of the parameter quantities of each model during training. The results show that the parameter quantities of the proposed model is significantly lower than that of other comparative models. The core reason for this advantage lies in the fact that, although all models adopt the same feature extractor architecture, the proposed model substantially reduces the parameter quantities of the classifier module by introducing a graph-structured optimization design. Notably, the CAN model has an extremely large parameter quantities, which is attributed to its architectural design: it employs two feature extractors (a source domain feature extractor and a target domain feature extractor) and three classifiers (a source domain classifier, a labeled target domain classifier, and an unlabeled target domain classifier), with no parameter sharing among all feature extractors and classifiers, leading to a significant increase in parameter scale.

In Fig. 3, we further statistically analyze the average time consumption of each model to complete one training epoch on the PPMI dataset. Among them, the single-epoch training time of the proposed model is only 83.78 seconds, while the training time of other models is higher than this value. Particularly, the DeCoTa model exhibits a remarkably long

Table 1. Hyperparameter analysis on three datasets

Hyperparameter	Dataset	Setting	ACC	SEN	F1	AUC
K	ADNI	10	0.862±0.11	0.858±0.11	0.857±0.11	0.961±0.03
		15	0.866±0.01	0.861±0.01	0.859±0.01	0.962±0.01
		20	0.863±0.12	0.859±0.12	0.857±0.12	0.961±0.03
		25	0.853±0.06	0.848±0.06	0.846±0.07	0.962±0.03
	AIBL	10	0.864±0.07	0.861±0.07	0.861±0.06	0.957±0.05
		15	0.877±0.05	0.874±0.05	0.872±0.06	0.962±0.02
		20	0.862±0.06	0.859±0.06	0.857±0.06	0.958±0.02
		25	0.848±0.12	0.845±0.11	0.843±0.10	0.943±0.05
	PPMI	10	0.889±0.02	0.744±0.03	0.852±0.02	1.000±0.01
		15	0.833±0.02	0.615±0.03	0.762±0.03	0.996±0.04
		20	0.822±0.03	0.590±0.07	0.739±0.06	1.000±0.00
		25	0.733±0.04	0.385±0.09	0.553±0.08	0.998±0.04
α	ADNI	0	0.837±0.06	0.831±0.05	0.826±0.05	0.954±0.02
		0.2	0.866±0.01	0.861±0.01	0.859±0.01	0.962±0.01
		0.5	0.856±0.13	0.851±0.14	0.850±0.13	0.959±0.03
		0.8	0.807±0.13	0.801±0.13	0.798±0.14	0.941±0.03
	AIBL	0	0.747±0.06	0.756±0.06	0.735±0.08	0.915±0.02
		0.2	0.877±0.05	0.874±0.05	0.872±0.06	0.962±0.02
		0.5	0.856±0.04	0.852±0.04	0.848±0.05	0.949±0.14
		0.8	0.822±0.10	0.515±0.10	0.801±0.14	0.941±0.02
	PPMI	0	0.757±0.02	0.462±0.05	0.632±0.05	0.964±0.01
		0.2	0.889±0.02	0.744±0.03	0.852±0.02	1.000±0.01
		0.5	0.844±0.04	0.641±0.09	0.779±0.07	0.999±0.04
		0.8	0.727±0.04	0.369±0.09	0.538±0.11	0.996±0.07

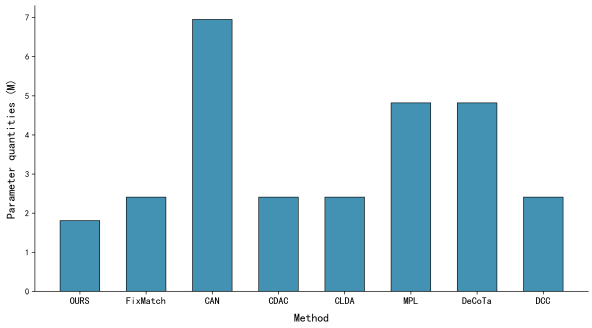


Figure 2. Parameter Quantities Comparison of Different Algorithms

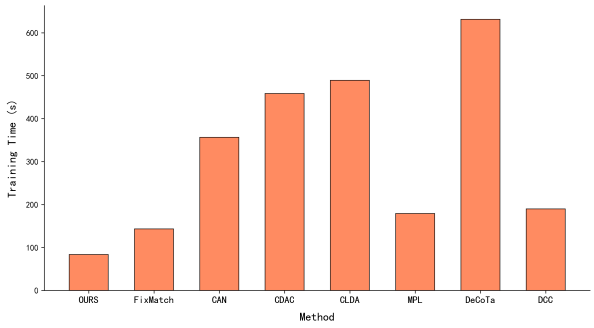


Figure 3. Training Time Comparison of Different Algorithms

single-epoch training time of 631.45 seconds. This phenomenon is presumably due to its technical design: the model decomposes the semi-supervised domain adaptation task into two subtasks (semi-supervised learning and unsupervised domain adaptation) and adopts a co-training mechanism. The parallel processing of dual tasks results in a significant increase in training time.

The above experimental analysis indicates that, compared with other models, the proposed model maintains a shorter training time while having a smaller parameter quantities, which fully demonstrates the significant advantage of DSDA in terms of training overhead.

Table 2. Comparison of diagnostic performance of different methods on PPMI dataset for Parkinson's disease

Method	Published	Dataset	ACC	SEN	F1	AUC
FixMatch[5]	2020 NIPS	PPMI	0.757±0.10	0.426±0.22	0.567±0.23	0.927±0.01
MPL[3]	2021 CVPR	PPMI	0.870±0.05	0.700±0.13	0.816±0.10	0.877±0.02
CDAC[2]	2021 CVPR	PPMI	0.870±0.05	0.701±0.12	0.819±0.08	0.949±0.01
CLDA[4]	2021 NIPS	PPMI	0.807±0.01	0.554±0.03	0.712±0.02	0.954±0.03
DeCoTa[8]	2021 ICCV	PPMI	0.823±0.02	0.592±0.03	0.743±0.03	0.983±0.01
ODADA[6]	2022 MIA	PPMI	0.627±0.07	**1.000±0.00**	0.632±0.04	0.972±0.05
DCC[1]	2023 ICME	PPMI	0.647±0.07	0.831±0.17	0.577±0.10	0.847±0.03
SLA[9]	2023 CVPR	PPMI	0.833±0.00	**1.000±0.00**	0.839±0.00	**1.000±0.00**
DDSPSeg[11]	2024 MIA	PPMI	0.823±0.06	0.592±0.14	0.735±0.11	0.932±0.01
CAN[7]	2025 TIM	PPMI	0.789±0.10	0.923±0.09	0.707±0.06	0.857±0.06
FSSADA[10]	2025 WCNC	PPMI	0.757±0.13	0.801±0.20	0.714±0.12	**1.000±0.00**
DSDA(ours)		PPMI	**0.889±0.02**	0.744±0.03	**0.852±0.02**	**1.000±0.01**

Table 3. Ablation Studies on the ADNI Dataset

	Method	ACC	SEN	F1	AUC	L_{label}	L_{con}	L_{NTA}	L_{pseudo}
1	Baseline	0.837±0.06	0.831±0.05	0.826±0.05	0.954±0.02	✓			
2	Baseline+PNIA	0.852±0.08	0.847±0.08	0.854±0.08	0.959±0.02	✓	✓		
3	Baseline+PNIA+NTA	0.862±0.04	0.858±0.05	0.856±0.03	0.961±0.02	✓	✓	✓	
4	Baseline+PNIA+CHCF	0.860±0.04	0.857±0.04	0.856±0.04	0.958±0.04	✓	✓		✓
5	Baseline+PNIA+NTA+CHCF	0.866±0.01	0.861±0.01	0.859±0.01	0.962±0.01	✓	✓	✓	✓

5. Comparison with Competing Methods on the PPMI Dataset

The experimental results on the PPMI dataset are shown in Tab. 2, our proposed DSDA method outperformed existing state-of-the-art approaches across ACC, F1, and AUC metrics, achieving 0.889, 0.852, and 1.000 respectively. These results not only demonstrate the method's robustness on this specific task but also validate the broader effectiveness of the DSDA model in other brain disorder diagnostic scenarios beyond its initial training context, underscoring its generalizability across different neurodegenerative conditions.

It is also worth noting that some methods report perfect scores (1.000) for SEN and AUC. Upon closer investigation, this anomaly is attributed to the limited number of available PET samples in the PPMI dataset. Such small-sample characteristics may lead to evaluation metric saturation or instability, producing extreme or misleading values. This finding highlights the need for robust evaluation strategies in future research, especially when working with small datasets, to mitigate the risk of evaluation bias and ensure the reliability of experimental conclusions.

6. Ablation Study

We also conducted the same ablation experiments on the ADNI and PPMI datasets, with the results as shown in Tabs. 3 and 4. It can be observed from the results that as different functional modules are gradually integrated into the model, the disorder prediction accuracy exhibits a sustained upward trend. This phenomenon fully demonstrates that each module we designed possesses independent and effective functions, making substantive contributions to the improvement of model performance. Notably, the effectiveness of the modules exhibits dataset dependence: on the ADNI dataset, adding only the NTA module (i.e., row 3) outperforms adding only the CHCF module (i.e., row 4), while the opposite holds true for the other datasets. This may be attributed to the fact that we address the two core issues—domain discrepancy elimination and discriminability preservation—by adopting a differential processing mechanism. Specifically, the NTA module focuses on domain discrepancy elimination, while the CHCF module focuses on discriminability preservation. Due to inherent differences in the degree of importance of these two issues across different datasets, the magnitude of model performance improvement naturally exhibits differentiation when optimizing for only one of the issues.

Table 4. Ablation Studies on the PPMI Dataset

	Method	ACC	SEN	F1	AUC	L_{label}	L_{con}	L_{NTA}	L_{pseudo}
1	Baseline	0.757±0.02	0.462±0.05	0.632±0.05	0.964±0.01	✓			
2	Baseline+PNIA	0.839±0.05	0.628±0.11	0.767±0.09	0.982±0.01	✓	✓		
3	Baseline+PNIA+NTA	0.870±0.02	0.854±0.02	0.864±0.02	0.993±0.05	✓	✓	✓	
4	Baseline+PNIA+CHCF	0.873±0.02	0.864±0.02	0.854±0.02	0.991±0.01	✓	✓		✓
5	Baseline+PNIA+NTA+CHCF	0.889±0.02	0.744±0.03	0.852±0.02	1.000±0.01	✓	✓	✓	✓

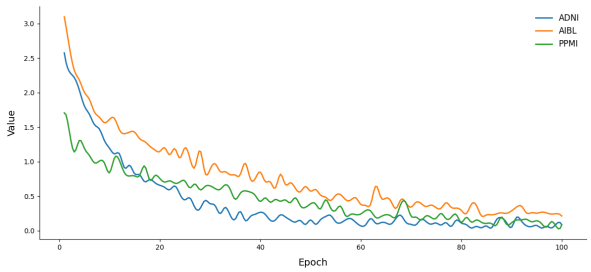


Figure 4. Visualization curves of the loss function on three datasets

7. Loss Visualization

The overall loss of the model on the three datasets is visualized, with the results presented in Fig. 4. The model was trained for a total of 100 epochs, and as can be observed from the figure, the overall loss on all three datasets continuously decreased during training: there was a significant drop in the early training stage, and despite minor fluctuations throughout the process, the overall downward trend was maintained. After the 80th training epoch, the three loss curves all gradually stabilized without noticeable decline or severe fluctuations, indicating that the model has fully converged.

8. Statistical Significance Tests

We conducted statistical significance tests on the ADNI, AIBL, and PPMI datasets. The results are shown in Tab. 5. Our method achieves the 'a' label on all datasets and outperforms other competing methods, demonstrating its statistically significant performance. In each row, different letters indicate significant differences ($P < 0.05$); same letters indicate no significant differences($P > 0.05$).

References

[1] Yidan Fan, Wenhuan Lu, and Yahong Han. Discriminative and contrastive consistency for semi-supervised domain adaptive image classification. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1074–1079, 2023. 2, 4

[2] Jichang Li, Guanbin Li, Yemin Shi, and Yizhou Yu. Cross-domain adaptive clustering for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2505–2514, 2021. 2, 4

[3] Hieu Pham, Zihang Dai, Qizhe Xie, and Quoc V. Le. Meta pseudo labels. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11552–11563, 2021. 2, 4

[4] Ankit Singh. Clda: Contrastive learning for semi-supervised domain adaptation. *Advances in neural information processing systems*, 34:5089–5101, 2021. 2, 4

[5] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020. 2, 4

[6] Y. Sun, D. Dai, and S. Xu. Rethinking adversarial domain adaptation: Orthogonal decomposition for unsupervised domain adaptation in medical image segmentation. *Medical Image Analysis*, 82:102623, 2022. 4

[7] Yiru Wang, Peng Wu, Senlin Fang, Bozhan Cao, Xinyu Wu, Xu Lu, and Zhengkun Yi. Semisupervised domain adaptation for wafer map defect recognition via cross-alignment network. *IEEE Transactions on Instrumentation and Measurement*, 74:1–13, 2025. 2, 4

[8] Luyu Yang, Yan Wang, Mingfei Gao, Abhinav Shrivastava, Kilian Q. Weinberger, Wei-Lun Chao, and Ser-Nam Lim. Deep co-training with task decomposition for semi-supervised domain adaptation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8886–8896, 2021. 2, 4

[9] Yu-Chu Yu and Hsuan-Tien Lin. Semi-supervised domain adaptation with source label adaptation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 24100–24109, 2023. 4

[10] Tongli Zeng, Shuo Chang, Jiashuo He, Shun Xu, Zhoushi Zhao, Sai Huang, and Zhiyong Feng. Towards cross-channel scenarios: Fusion semi-supervised adversarial domain adaptation modulation classification network. In *2025 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2025. 4

[11] B. Zheng, R. Zhang, S. Diao, J. Zhu, Y. Yuan, J. Cai, L. Shao, S. Li, and W. Qin. Dual domain distribution disruption with semantics preservation: Unsupervised domain adaptation for medical image segmentation. *Medical Image Analysis*, 97: 103275, 2024. 4

Table 5. Statistical Significance Test

Dataset	FixMatch	MPL	CDAC	CLDA	DeCoTa	ODADA	DCC	SLA	DDSPSeg	CAN	FSSADA	OURS
ADNI	0.741d	0.714e	0.775d	0.753d	0.837bc	0.655f	0.569g	0.819c	0.853ab	0.452h	0.652f	0.866a
AIBL	0.793c	0.690e	0.653f	0.822b	0.750d	0.763d	0.800c	0.563g	0.842b	0.462h	0.683e	0.877a
PPMI	0.757f	0.870b	0.627h	0.823c	0.870b	0.807d	0.823c	0.647g	0.833c	0.789d	0.757f	0.889a