

Topology-Driven Transferability Estimation of Medical Foundation Models for Segmentation

Jiaqi Tang^{1,4,5,6*}, Shaoyang Zhang^{2*}, Xiaoqi Wang³, Jiaying Zhou¹, Yang Liu¹, and Qingchao Chen^{1,4,5,6}✉

¹ Peking University, Beijing, China

² Hohai University, Nanjing, China

³ Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China

⁴ National Institute of Health Data Science, Peking University, Beijing, China

⁵ Institute of Medical Technology, Peking University, Beijing, China

⁶ State Key Laboratory of General Artificial Intelligence, Peking University
{jiaqi_tang, qingchao.chen}@pku.edu.cn, zsyhhu04@163.com

✉ Qingchao Chen is the corresponding author

Abstract. The advent of large-scale self-supervised learning (SSL) has produced a vast zoo of medical vision foundation models. However, selecting optimal one for specific segmentation tasks remains a computational bottleneck. Existing Transferability Estimation (TE) metrics, primarily designed for classification, rely on global statistical assumptions and fail to capture the topological complexity essential for dense prediction. We propose a novel Topology-Driven Transferability Estimation framework that evaluates manifold tractability rather than statistical overlap. Our approach introduces three components: (1) Global Representation Topology Divergence (GRTD), utilizing Minimum Spanning Trees to quantify feature-label structural isomorphism; (2) Local Boundary-Aware Topological Consistency (LBTC), which assesses manifold separability specifically at critical anatomical boundaries; and (3) Task-Adaptive Fusion, which dynamically integrates global and local metrics based on the semantic cardinality of the target task. Validated on the large-scale OpenMind benchmark across diverse anatomical targets and SSL-based vision foundation models, our approach significantly outperforms state-of-the-art baselines by around **31%** relative improvement in the weighted Kendall’s τ , providing a robust, training-free proxy for efficient model selection without the cost of fine-tuning. The code is available at <https://github.com/SPIresearch/TopoTE>.

Keywords: transferability estimation · transfer learning · model selection

1 Introduction

The rise of large-scale self-supervised learning (SSL) has created a diverse zoo of medical vision foundation models trained on massive unlabeled data to learn

* These authors contributed equally to this work.

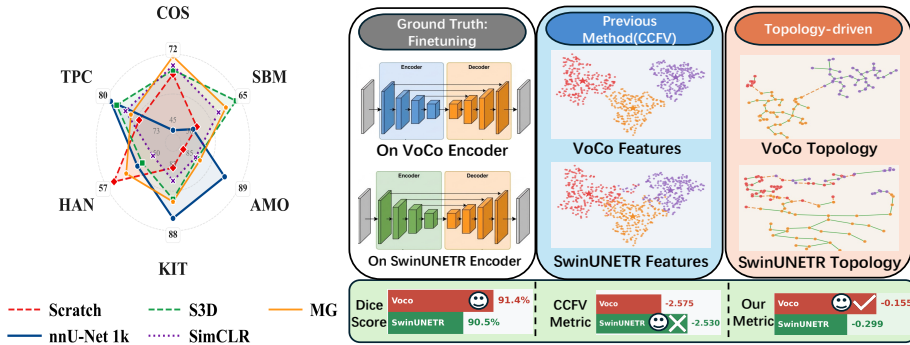


Fig. 1: **Left:** Fine-tuning performance of diverse vision foundation models varies across downstream datasets, indicating that the optimal pre-trained encoder is task-dependent. **Right:** Visualizing transferability: Statistics vs. Topology. We compare the ground truth fine-tuning performance (Left) against rankings from the statistical metric (Middle) and our topology-driven metric (Right).

general-purpose representations [2, 3, 6, 16, 21, 25]. Yet, the best pre-trained encoder is highly task-dependent: as illustrated in Fig. 1 (left), different downstream segmentation datasets favor different source models, making exhaustive fine-tuning a costly combinatorial search[20]. This motivates a training-free Transferability Estimation (TE) framework that predicts post-fine-tuning segmentation performance directly from pre-training features, enabling fast and resource-efficient model selection.

Existing TE metrics, largely developed for image classification [4, 13–15, 23, 24], can be misaligned with the “SSL encoder + segmentation decoder” setting. Methods such as LEEP [13] and LogME [24] implicitly favor linear separability, while embedding-based scores such as GBC [14] and CCFV [23] often assume simple parametric statistics (e.g., Gaussianity). However, segmentation quality depends less on coarse global class separation and more on whether features preserve local geometric structure near high-frequency boundaries. As shown in Fig. 1 (right), purely statistical similarity may yield rankings inconsistent with fine-tuning outcomes, whereas a topology-aware view of the feature manifold better reflects boundary separability and thus transferability.

Our method bridges these gaps through a topology-driven approach. **Rather than forcing complex feature spaces into predefined statistical molds, we utilize non-parametric, graph-theoretic structures to quantify manifold alignment.** By explicitly modeling local boundary regions and dynamically balancing global and local priors, our framework could match the specific structural complexity of diverse downstream medical targets. Specifically, our method introduces three novel components tailored for medical segmentation, as shown in Figure 2: (1) *Representation Topology Divergence (GRTD)* quantifies structural alignment by measuring the discrepancy between Minimum Spanning Trees (MST) constructed in the feature and label spaces, capturing the over-

all manifold isomorphism. (2) *Local Boundary-Aware Topological Consistency (LBTC)* explicitly assesses manifold separability at critical boundary patches, utilizing local MST graphs to ensure distinct decision boundaries in the presence of background heterogeneity, where segmentation typically fails. (3) *Task-Adaptive Topological Fusion* dynamically calibrates the global and local metrics based on the target task’s complexity, optimally balancing the need for broad anatomical context versus fine-grained boundary details.

We validate our framework on the large-scale OpenMind benchmark [20], covering 6 diverse anatomical segmentation tasks and a model zoo consisting of 7 mainstream SSL methods pre-trained on 114,000 3D volumes. Extensive experiments demonstrate that our topology-driven approach significantly outperforms existing baselines by around **31%** relative improvement in weighted Kendall’s τ , offering a robust, source-free solution for efficient model selection in the era of medical vision foundation models.

2 Methodology

2.1 Problem Formulation and Framework Overview

We address the problem of selecting the optimal SSL-based vision foundation model from the model zoo $\mathcal{M} = \{\phi_k\}_{k=1}^K$ for a target segmentation task $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$. Let $P(\phi, \mathcal{D})$ denote the ground-truth performance (e.g., Dice score) after fine-tuning. Our objective is to derive a *training-free* transferability score $T(\phi, \mathcal{D})$ that serves as a reliable proxy for P . Note that “training-free” denotes requiring zero gradient updates and operating in an inference-only way. Ideally, the score should preserve the performance ranking across the model zoo, satisfying the monotonicity condition for any pair of models $\phi_i, \phi_j \in \mathcal{M}$:

$$T(\phi_i, \mathcal{D}) > T(\phi_j, \mathcal{D}) \iff P(\phi_i, \mathcal{D}) > P(\phi_j, \mathcal{D}) \quad (1)$$

While existing metrics typically rely on distributional statistics and neglect the complex geometry of SSL manifolds required for dense prediction, we argue that for segmentation tasks, *topological tractability*, i.e., the preservation of structural connectivity or boundary separation, is a rather reliable predictor of transferability than mere statistical overlap. We propose a topology-driven framework that assesses feature-label alignment at three levels: (1) **Global Representation Topology Divergence (GRTD)**, which quantifies the structural discrepancy between feature-induced and label-induced Minimum Spanning Trees (MST); (2) **Local Boundary-Aware Topological Consistency (LBTC)**, which specifically examines the feature space distinctness at anatomical boundaries, ensuring that the pre-trained features remain separable in these critical transition zones where segmentation failures most frequently occur; and (3) **Task-Adaptive Aggregation** strategy that dynamically weights global and local metrics based on the inherent topological complexity of the target anatomy.

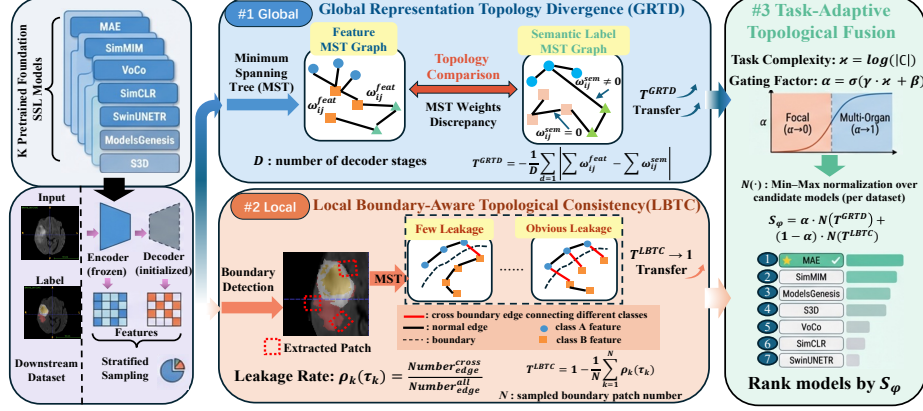


Fig. 2: The overall framework. First, we extract multi-scale features via stratified sampling and construct topological graphs via MST. We then compute GRTD to quantify overall manifold alignment and LBTC to evaluate separability at critical anatomical boundaries. These metrics are integrated via a task-complexity gating factor, producing a final score S_ϕ that predicts fine-tuning performance without training.

2.2 Global Representation Topology Divergence (GRTD)

To capture the complex, non-linear geometry without imposing parametric assumptions (e.g., Gaussianity), we model the data distribution using the Minimum Spanning Tree (MST). The MST serves as a robust non-parametric descriptor of the manifold’s 1-skeleton, enabling us to evaluate transferability through the lens of topological isomorphism: a highly transferable model should yield a feature topology that naturally aligns with the semantic hierarchy.

Let $\mathcal{X} = \{(\mathbf{v}_i, y_i)\}_{i=1}^N$ be the sampled feature-label pairs. We construct two graphs to evaluate topological alignment, as shown in Figure 2. The first is the *Native Feature Graph* $\mathcal{G}_{feat}$, where edge weights represent Euclidean distances in the embedding space: $w_{ij}^{feat} = \|\mathbf{v}_i - \mathbf{v}_j\|_2$. The second is the *Semantic Label-Induced Graph* $\mathcal{G}_{sem}$, which represents an ideal label-driven topology. It explicitly injects semantic constraints by forcing samples of the same class to perfectly cluster while preserving the original feature distances for inter-class pairs up to a maximum penalty:

$$w_{ij}^{sem} = \begin{cases} 0 & \text{if } y_i = y_j, \\ \min(\|\mathbf{v}_i - \mathbf{v}_j\|_2, \lambda) & \text{if } y_i \neq y_j. \end{cases} \quad (2)$$

where λ penalizes inter-class transitions. The MSTs derived from these graphs, denoted $\mathcal{T}_{feat}$ and $\mathcal{T}_{sem}$, represent the natural clustering tendency and the ideal semantic connectivity, respectively.

We define the GRTD as the discrepancy between the total weights of these two topological structures. Aggregating across D decoder stages, the final score

is formulated as:

$$T^{\text{GRTD}} = -\frac{1}{D} \sum_{d=1}^D \left| \sum_{(i,j) \in \mathcal{T}_{feat}^d} w_{ij}^{feat} - \sum_{(i,j) \in \mathcal{T}_{sem}^d} w_{ij}^{sem} \right|. \quad (3)$$

A higher T^{GRTD} (closer to 0) indicates that the encoder’s native geometry naturally respects semantic boundaries.

2.3 Local Boundary-Aware Topological Consistency (LBTC)

Although GRTD captures the global manifold topology, it is susceptible to the class imbalance typical of medical images, where low-frequency background regions dominate the feature space statistics. However, the success of segmentation often hinges on the model’s ability to preserve high-frequency details at critical anatomical boundaries. Hence, we propose the Local Boundary-Aware Topological Consistency (LBTC), shifting the focus from global isomorphism to local separability, explicitly scrutinizing whether the pre-trained features maintain distinct decision boundaries within ambiguous boundary patches.

We first define the set of boundary anchors $\partial\mathcal{Y}$ via the morphological gradient of the ground truth masks. For each anchor $c_k \in \partial\mathcal{Y}$, we extract a local patch $\mathcal{P}_k$ and construct a local graph $\mathcal{G}_k$. The Minimum Spanning Tree of this local neighborhood, $\mathcal{T}_k$, serves as a probe for local cluster purity. Ideally, $\mathcal{T}_k$ should traverse all intra-class nodes before bridging the semantic gap. We quantify the *Topological Leakage Rate* ρ_k , which measures the proportion of edges in the local MST that erroneously connect distinct semantic classes:

$$\rho_k(\mathcal{T}_k) = \frac{1}{|\mathcal{V}_k| - 1} \sum_{(u,v) \in \mathcal{E}(\mathcal{T}_k)} \mathbb{I}(y_u \neq y_v), \quad (4)$$

where $\mathcal{E}(\mathcal{T}_k)$ denotes the edge set of the local MST and $\mathbb{I}(\cdot)$ is the indicator function. The final LBTC score is derived by aggregating these local inconsistencies over N sampled boundary regions, formulated as the complement of the expected leakage:

$$T^{\text{LBTC}} = 1 - \frac{1}{N} \sum_{k=1}^N \rho_k(\mathcal{T}_k). \quad (5)$$

A vanishing leakage rate indicates perfect local k-NN separability, and a score approaching 1 implies that the encoder preserves strict topological separation even within the ambiguous transition zones of the manifold.

2.4 Task-Adaptive Topological Fusion

Medical segmentation targets exhibit distinct topological regimes: multi-organ tasks demand global structural preservation (high structural complexity), while

Table 1: Weighted Kendall’s τ for transferability estimation on the OpenMind benchmark.

Method	ID (Same Region)				OOD (Different Region)		Avg
	MSF	ISL	HNT	TPC	ACD	KIT	
LogME [24]	−0.524	−0.223	−0.315	−0.545	0.575	−0.649	−0.280
LEEP [13]	−0.220	−0.427	−0.492	−0.051	−0.621	0.225	−0.264
GBC [14]	−0.707	−0.601	−0.537	−0.256	−0.325	−0.621	−0.508
CCFV [23]	0.277	0.503	0.869	0.664	0.817	0.180	0.552
Ours	0.705	0.814	0.905	0.664	0.671	0.578	0.723

small lesion extraction relies on local boundary contrast (high boundary complexity). A static combination of global and local metrics fails to generalize across these heterogeneous distributions. To resolve this, we propose a dynamic fusion mechanism driven by the semantic cardinality of the target task.

We define a task complexity prior $\kappa = \log(|\mathcal{C}|)$, where $|\mathcal{C}|$ is the number of semantic classes. This prior modulates a gating factor $\alpha \in (0, 1)$ via a sigmoid function $\sigma(\cdot)$, shifting the focus between macro-structure and micro-details. Let $\mathcal{N}(\cdot)$ denote the Min-Max normalization operator over the model zoo. The final transferability score $\mathcal{S}_\phi$ is computed as a convex combination of the global and local topological metrics:

$$\alpha = \sigma(\gamma \cdot \kappa + \beta), \quad \mathcal{S}_\phi = \alpha \cdot \mathcal{N}(T^{\text{GRD}}) + (1 - \alpha) \cdot \mathcal{N}(T^{\text{LBT}}), \quad (6)$$

where γ and β are scaling constants. This formulation ensures that for complex anatomical tasks ($\alpha \rightarrow 1$), the ranking prioritizes global layout isomorphism, whereas for focal pathologies ($\alpha \rightarrow 0$), it emphasizes the sharpness of local decision boundaries.

3 Experiment and Results

3.1 The Foundation Model Zoo and Downstream Tasks

We utilize the OpenMind benchmark [20] model zoo, featuring ResEnc-L [11] backbones pre-trained on 114,000 unlabeled 3D brain MRIs. To rigorously evaluate transferability without prior label exposure, we select models trained via reconstruction (MAE[6], SimMIM[3], ModelsGenesis[25], S3D [19]) and contrastive learning (VoCo[21], SimCLR[2], SwinUNETR[16]). We evaluate these models across diverse downstream tasks spanning various anatomies and modalities.

In-Distribution (ID) Tasks: We select four datasets closely aligned with the pre-training domain (head-and-neck) to test intrinsic transferability: *ISLES* (*ISL*) [9] (stroke lesions on DWI/ADC), *HNTS-MRG* (*HNT*) [18] (head-and-neck tumors on MR), *MS FLAIR* (*MSF*) [12] (multiple-sclerosis lesions), and

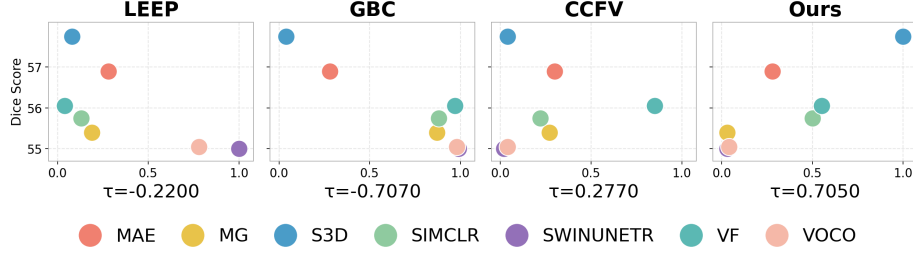


Fig. 3: Correlation between the fine-tuning performance and transferability metrics using MSF as an example. The vertical axis represents the average Dice, while the horizontal axis represents the *standardized* transferability metric. We want to observe a positive relationship between higher performance and higher transferability estimations.

ToP-CoW (TPC) [22] (Circle of Willis vessels). **Out-of-Distribution (OOD) Tasks:** To assess generalization under distribution shifts, we include two OOD tasks: *ACDC (ACD)* [1] (cardiac structures on cine-MRI), and *KiTS19 (KIT)* [8] (kidney and tumors on CT). Notably, KIT represents a challenging cross-anatomy and cross-modality (MR→CT) transfer scenario with clear tumor boundaries [11].

3.2 Implementation Details

Considering some SSL-based candidate models just have a well-trained encoder, we attach a randomly initialized nnU-Net [10] segmentation decoder to these models, extracting multi-scale features via sliding-window inference. To ensure topological fidelity despite severe class imbalance, we employ a stratified sampling strategy that balances class-conditional foreground voxels against background and global distributions. We differentiate the extraction depth to align with topological roles: GRTD is computed on the final k decoder layers to capture semantic layout, while Local LBTC is evaluated on the last k encoder layers to assess intrinsic boundary separability. The task-adaptive gating weights are estimated using a computationally efficient pilot set of $K=10$ cases, with final rankings quantified via the weighted Kendall correlation ($\tau^{*,w}$) [17].

3.3 Results and Analysis

Overall Performance We report the weighted Kendall’s τ between the estimated transferability scores and the actual fine-tuning performance. As shown in Table 1 and Figure 3, traditional classification-based metrics (LogME, LEEP) exhibit severe negative correlations, confirming their inadequacy for dense prediction tasks. While the embedding-based CCFV shows positive correlations, it struggles on challenging out-of-distribution (OOD) transfers like KIT ($\tau = 0.180$). In contrast, our method demonstrates robust and consistent performance across both in-distribution and OOD scenarios, achieving the highest average

Table 2: **Left:** weighted Kendall’s τ_w^* averaged over five random seeds for each decoder initialization scheme. **Right:** transferability time and fine-tuning time are averaged over five repeated timing trials; both are reported as the *total* wall-clock time for all 7 models. *All times are in minutes.*

Decoder initialization robustness				Time comparison			
Dataset	Kaiming	Xavier	Gaussian	Dataset	CCFV	Ours	Fine-tune
TPC	0.664	0.646	0.652	TPC	1050.5	10.07	3000+
ACD	0.671	0.624	0.652	ACD	756.9	8.02	3000+
MSF	0.705	0.706	0.718	MSF	22.3	2.89	3000+
Avg.	0.680	0.659	0.674	Avg.	609.9	6.99	3000+

Table 3: Ablation study on topological priors of different datasets. We stratify tasks into *Fragmented* (ISL, MSF) and *Structured* (ACD, TPC) targets. Our **Adaptive** fusion achieves robust performance across all regimes.

Method	ISL	MSF	ACD	TPC	Avg.
Global (GRTD)	0.122	0.333	0.624	0.664	0.436
Local (LBTC)	0.814	0.705	0.319	0.062	0.475
Adaptive (Ours)	0.814	0.705	0.671	0.664	0.714

correlation of 0.723, validating that topological tractability is a superior proxy for segmentation transferability.

Robustness to Decoder Initialization. For candidate models requiring a randomly initialized decoder, our metric should reflect the intrinsic quality of the *pre-trained* representation rather than stochasticity from initialization. As summarized in Table 2 (left), the weighted Kendall’s τ_w^* remains consistently stable across Kaiming [7]/Xavier [5]/Gaussian initializations with small variance, suggesting that the topology-driven signal dominates over initialization noise.

Computational Time Comparison. Table 2 (right) contrasts the practical cost of model selection via training-free scoring versus exhaustive downstream fine-tuning. Compared with CCFV, our approach substantially reduces the metric-computation overhead, enabling rapid screening over a model zoo. More importantly, it avoids the prohibitive end-to-end cost of repeating fine-tuning across multiple candidates, making transferability estimation a realistic substitute for brute-force selection in resource-constrained settings.

Impact of Target Topology Table 3 reveals the complementary nature of our topological priors. While Global (GRTD) and Local (LBTC) metrics suffer significant performance drops on fragmented and structured targets respectively, our adaptive fusion effectively bridges this gap and achieves a robust average of **0.714**, a substantial improvement over the single-stream baselines.

4 Conclusion

In this paper, we present a topology-driven transferability estimation framework of medical vision foundation models for segmentation. By shifting the paradigm from statistical overlap to manifold topology, our approach captures both global structural isomorphism and local boundary separability. Extensive validation on the OpenMind benchmark demonstrates that our task-adaptive method significantly outperforms existing baselines across diverse tasks. By providing a robust, training-free proxy for model selection, our framework eliminates the prohibitive computational costs of exhaustive fine-tuning, paving the way for the efficient and scalable clinical deployment of medical foundation models. Future work will focus on extending our framework to promptable foundation models, such as the MedSAM series, by incorporating prompt-conditioned topological priors. Furthermore, we plan to investigate the applicability of our metric across a wider spectrum of architectures and training paradigms, including ViT-based models and DINO-style self-distillation frameworks, to further establish the universality of topology-driven transferability estimation.

Acknowledgments. This work was supported by grants from the National Natural Science Foundation of China (62571009). The Civil Space Technology Pre-research Foundation NO.D010101.

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

References

1. Bernard, O., Lalonde, A., Zotti, C., Cervénanský, J., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging* (2018)
2. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: *International conference on machine learning*. pp. 1597–1607. PmLR (2020)
3. Chen, Z., Agarwal, D., Aggarwal, K., Safta, W., Balan, M.M., Brown, K.: Masked image modeling advances 3d medical image analysis. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 1970–1980 (2023)
4. Ding, N., Chen, X., Levinboim, T., Changpinyo, S., Soricut, R.: Pactran: Pac-bayesian metrics for estimating the transferability of pretrained models to classification tasks. In: *European Conference on Computer Vision*. pp. 252–268. Springer (2022)
5. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
6. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 16000–16009 (June 2022)

7. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034 (2015)
8. Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., et al.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. *Medical image analysis* **67**, 101821 (2021)
9. Hernandez Petzsche, M.R., De La Rosa, E., Hanning, U., Wiest, R., Valenzuela, W., Reyes, M., Meyer, M., Liew, S.L., Kofler, F., Ezhov, I., et al.: Isles 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Scientific data* **9**(1), 762 (2022)
10. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2), 203–211 (2021)
11. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)
12. Muslim, A.M., Mashohor, S., Al Gawwam, G., Mahmud, R., Hanafi, M.B., Alnuaimi, O., Josephine, R., Almutairi, A.D.: Brain mri dataset of multiple sclerosis with consensus manual lesion segmentation and patient meta information. *Data in Brief* **42**, 108139 (2022). <https://doi.org/10.1016/j.dib.2022.108139>, <https://www.sciencedirect.com/science/article/pii/S235234092200347X>
13. Nguyen, C., Hassner, T., Seeger, M., Archambeau, C.: Leep: A new measure to evaluate transferability of learned representations. In: International Conference on Machine Learning. pp. 7294–7305. PMLR (2020)
14. Pándy, M., Agostinelli, A., Uijlings, J., Ferrari, V., Mensink, T.: Transferability estimation using bhattacharyya class separability. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9172–9182 (2022)
15. Tan, Y., Li, Y., Huang, S.L.: Otce: A transferability metric for cross-domain cross-task representations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 15779–15788 (2021)
16. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20730–20740 (June 2022)
17. Vigna, S.: A weighted correlation index for rankings with ties. In: Proceedings of the 24th international conference on World Wide Web. pp. 1166–1176 (2015)
18. Wahid, K., Dede, C., Naser, M., Fuller, C.: Training dataset for hntsmrg 2024 challenge. <https://zenodo.org/doi/10.5281/zenodo.11199559> (2024)
19. Wald, T., Ulrich, C., Lukyanenko, S., Goncharov, A., Paderno, A., Miller, M., Maerkisch, L., Jaeger, P., Maier-Hein, K.: Revisiting mae pre-training for 3d medical image segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5186–5196 (2025)
20. Wald, T., Ulrich, C., Suprijadi, J., Ziegler, S., Nohel, M., Peretzke, R., Kohler, G., Maier-Hein, K.: An openmind for 3d medical vision self-supervised learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 23839–23879 (2025)
21. Wu, L., Zhuang, J., Chen, H.: Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis. In: Proceedings of the IEEE/CVF

- Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22873–22882 (June 2024)
22. Yang, K., Musio, F., Ma, Y., Juchler, N.: Topcow: Benchmarking topology-aware anatomical segmentation of the circle of willis (cow) for cta and mra. *Journal of Medical Imaging Technology* **34**, 123–135 (2024)
 23. Yang, Y., Wei, M., He, J., Yang, J., Ye, J., Gu, Y.: Pick the best pre-trained model: Towards transferability estimation for medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 674–683. Springer (2023)
 24. You, K., Liu, Y., Wang, J., Long, M.: Logme: Practical assessment of pre-trained models for transfer learning. In: *International Conference on Machine Learning*. pp. 12133–12143. PMLR (2021)
 25. Zhou, Z., Sodha, V., Pang, J., Gotway, M.B., Liang, J.: Models genesis. *Medical image analysis* **67**, 101840 (2021)