BioGuide: Biomedically-Guided Segmentation for Medical Tasks

Nadezhda Alsahanova[1,2][0009−0002−8185−2493], Yauhen Statsenko[3][0000−0002−7713−3333], Alexander Bernstein[1][0000−0002−5250−1849], Rifat Hamoudi[2,4,5][0000−0002−1402−0868], and Maxim Sharaev[1,2,6][0000−0002−5670−2891]

[1] Applied AI Institute, n.alsahanova@applied-ai.ru
[2] BIMAI-Lab, University of Sharjah, Sharjah, United Arab Emirates
[3] College of Medicine and Health Sciences, UAEU, United Arab Emirates
[4] Research Institute of Medical and Health Sciences, University of Sharjah, United Arab Emirates
[5] Division of Surgery and Interventional Science, University College London, London, United Kingdom
[6] Translational Oncology Research Center, QBRI, College of Health and Life Sciences, HBKU, Qatar Foundation, Doha, Qatar

Abstract. While deep learning models excel at medical segmentation, they typically do not leverage structured clinical descriptors that could inform feature learning. Here we show that integrating structured clinical descriptors into segmentation models guides learning and boosts performance. We introduce BioGuide, a lightweight multi-task learning framework that integrates fine-grained radiological descriptors as auxiliary targets to guide feature learning. A small MLP head is attached to the encoder bottleneck of standard segmentation networks, encouraging latent representations to align with clinically meaningful patterns, boosting potential interpretability. We evaluate BioGuide across four diverse 2D and 3D brain MRI datasets for Multiple Sclerosis (MS) and Focal Cortical Dysplasia (FCD) using state-of-the-art segmentation models. Our results show that incorporating biomedical guidance consistently improves segmentation scores. BioGuide yields a consistent improvement of up to 3.1% in Dice score and 8.5% in lesion detection rate across all tested datasets, demonstrating superior data efficiency. Code is available at https://github.com/NadezhdaAlsahanova/BioGuide.

Keywords: MRI Segmentation · Domain knowledge · Focal Cortical Dysplasia · Multiple Sclerosis.

1 Introduction

Precise segmentation of brain lesions is critical for a wide range of clinical applications [3, 9, 15]. However, the accuracy of automatic segmentation can be low due to phenotypic heterogeneity of diseases. Some phenotypes are unremarkable,

and they can be easily missed. For example, subtle foci of gray-white matter blurring in FCD [2] and small isolated juxtacortical lesions in MS [4]. The accurate detection and delineation of these lesions require novel approaches.

Despite recent advances, important limitations remain in current segmentation approaches. Commonly, deep learning models rely solely on voxel-level characteristics and often fail to capture clinically structured information. To bridge this gap researchers proposed Multi-Task Learning (MTL) which uses different approaches [8, 5]. Some MTL methods focus on high-level outcomes like survival or treatment response [1]. However, this information does not foster the accuracy of segmentation largely. Advanced MTL approaches rely on complex, task-specific architectures that are difficult to adapt to different pathologies or baseline models [12]. MTL models with broad labels, such as tumor grade, do not reach high performance due to the lack of a straight-forward assocication between the lables and radiological data [11]. Therefore, improving MTL requires both flexible, task-agnostic architectural designs and the explicit integration of radiological expertise to enhance segmentation accuracy and generalizability.

The primary aim of this study was to develop and validate a clinically guided MTL framework which integrates structured radiological descriptors into deep learning segmentation models. We hypothesized that incorporating biomedical domain knowledge as auxiliary targets would align latent feature representations with clinically meaningful patterns. To test this hypothesis, we introduced BioGuide. BioGuide integrates an auxiliary classification head—a simple Multi-Layer Perceptron (MLP)—positioned immediately after the encoder. In contrast to previous works, BioGuide utilizes fine-grained clinical descriptors as auxiliary targets. For FCD, these include specific radiological signs such as hyperintensity, blurring, and lobar location: frontal or temporal. For MS, the targets describe the typical anatomical distribution of lesions, including juxtacortical, subcortical, paraventricular and infratentorial regions. Lesion location correlates with lesion size and morphology. Therefore, incorporating anatomical descriptors may improve segmentation accuracy.

In this work, we evidence the generalizability of BioGuide framework by evaluating it across two distinct tasks of lesion segmentation in MS and FCD cases. We used diverse datasets to show that the models built within the designed framework consistently outperform the current SOTA baselines. These are the Transformer [14] and the nnU-Net [7]. Thus, our contribution is the following:

- We propose BioGuide, a novel framework that integrates clinical features (e.g., blurring, T2 hyperintensity) as auxiliary tasks, effectively regularizing existing segmentation models with domain-specific knowledge.
- The framework is flexible, as evidenced by its successful application to different architectures (nnU-Net and Transformers), various lesion types, and both 2D and 3D imaging data.
- We demonstrate through latent space visualization that BioGuide aligns features with clinical descriptors.

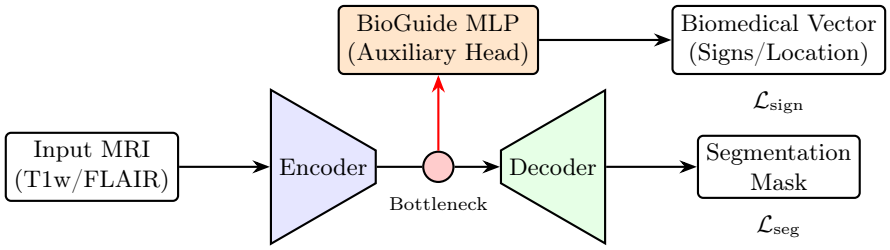

Fig. 1. BioGuide architecture with an auxiliary MLP head at the encoder bottleneck.

2 BioGuide: multi-task learning architecture

The BioGuide framework is a modular multi-task learning architecture designed to align deep learning feature extraction with established medical domain knowledge. The framework is designed to be "plug-and-play," allowing it to be integrated into any existing segmentation backbone (e.g., nnU-Net or Transformer) by attaching a lightweight auxiliary classification head to the encoder's bottleneck, see Fig. 1.

We represent clinical knowledge through a multi-label biomedical vector. For FCD, this vector encodes both intensity-based radiological signs (e.g., blurring, T2 hyperintensity, transmantle sign) and anatomical location (lobar indicators), see Fig. 2, resulting in a combined descriptor of the lesion's phenotype and position. For MS, the vector identifies the specific categories of lesions present in the image, see Fig. 2. In both cases, the vector is binary (1 for presence, 0 for absence). To ensure the model learns from relevant signs, if a specific 2D slice or 3D patch does not contain lesions, the vector is set to all zeros — equivalent to the standard no-object target in multi-label detection, since each descriptor encodes a property of a present lesion rather than the patch itself.

A lightweight MLP classification head acts as a regularizer, encouraging the latent space to become more linearly separable by phenotype, providing a strong global prior to the decoder. Because the descriptors are patient-level binary labels, the model is asked only whether a sign is present, not to localize it pixel-wise. Backpropagating this patch-level loss through the encoder forces early and mid-level filters to become sensitive to clinically relevant signs. While multi-scale guidance could be beneficial, we opted for this minimal and computationally efficient design to clearly isolate the effect of the clinical signal and maintain the 'plug-and-play' nature with any backbone.

The model is optimized using a joint loss function that balances the primary pixel-wise segmentation task with the auxiliary biomedical guidance task:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \lambda \mathcal{L}_{\text{sign}},$$

where $\mathcal{L}_{\text{seg}}$ is Dice-Tversky for segmentation and $\mathcal{L}_{\text{sign}}$ is Binary Cross-Entropy for sign prediction. A critical hyperparameter in our framework is the vector weight λ. A small weight λ ensures the auxiliary task provides inductive bias without distracting from segmentation.

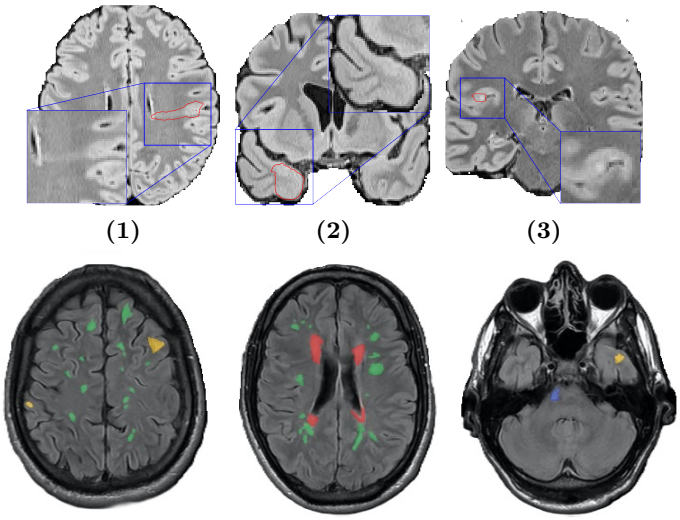

Fig. 2. Radiological signs of FCD and MS lesion distribution for BioGuide encoding. **Top:** FCDs on T1 MRI are shown with red contours indicating manual segmentations. Clinical features are encoded as 10D vectors $\mathbf{v}$: **(1)** parietal with blurring/transmantle sign, $\mathbf{v} = [1, 0, 0, 1, 0, 1, 0, 0, 0, 0]$; **(2)** temporal with blurring, $\mathbf{v} = [0, 0, 1, 1, 0, 0, 0, 0, 0, 0]$; **(3)** frontal with gray matter hyperintensity, $\mathbf{v} = [1, 0, 0, 0, 0, 0, 1, 0, 0, 0]$. **Bottom:** MS lesion subtypes across axial slices as 4D vectors $\mathbf{v}$ for Juxtacortical, Subcortical, Periventricular, and Infratentorial regions. **(1)**: $\mathbf{v} = [0, 1, 1, 0]$; **(2)**: $\mathbf{v} = [1, 1, 0, 0]$; **(3)**: $\mathbf{v} = [0, 0, 1, 1]$.

3 Experiments

3.1 Datasets and Data Preparation

Focal Cortical Dysplasia (FCD) We evaluate BioGuide on three FCD datasets utilizing $T1$-weighted and FLAIR modalities. Anonymized scans were registered to MNI152 space and skull-stripped via HD-BET [6]. Training was performed using 5-fold cross-validation. To ensure the integrity of the guidance signal, training subjects were selected to have only a single lesion with clear descriptive signs. Subjects with multiple lesions or ambiguous signs were reserved for independent testing (additional test sets). All lesions for internal datasets were manually segmented by an expert radiologist using ITK-SNAP [13].

Internal Dataset 1: 77 subjects (one fold of 16, others of 15) and a test set (T) of 11 subjects (1–12 lesions each). **Internal Dataset 2:** 79 subjects (one fold of 15, others of 16) and 7 additional test (T) subjects (1–3 lesions each). This dataset utilized lower-quality FLAIR sequences. **Open Dataset:** 77 subjects refined from the open dataset [10]. As the original metadata lacked structured signs, professional radiologist performed a retrospective review to label each case (8 cases were excluded due to ambiguous radiological signs).

Multiple Sclerosis (MS) To demonstrate generalizability, we utilized an internal MS dataset of 46 patients (1,102 2D FLAIR slices).Training was performed using 5-fold cross-validation. Images underwent intensity normalization. Unlike the textural focus of FCD, MS guidance focuses on anatomical location which dictates the expected shape and size of the lesions.

3.2 Baselines and Ablation Studies

For task-specific benchmarks, we compare our FCD results against the SOTA Multiscale Transformer [14] using its original architectural parameters. For the MS task, we utilize the nnU-Net [7], which serves as an automatically optimized U-Net baseline for medical image segmentation

To evaluate the impact of the BioGuide framework, we conducted ablation studies comparing several configurations. The **Baseline** consists of a standard segmentation network without the auxiliary head. To isolate specific domain knowledge contributions, we tested **Location-Only** and **Signs-Only** variants, which utilize only lobar positions or radiological signs as guidance, respectively. Random noise and shuffled labels served as **knowledge-blind** controls. These controls also serve as a degenerate lesion-present head: any non-zero shuffled vector still signals that a lesion exists. The **Full BioGuide** model incorporates both location and radiological sign prediction for comprehensive guidance.

3.3 Optimization and Hyperparameter Search

FCD Optimization Pipeline. To ensure a rigorous and unbiased evaluation, we performed a grid search independently for each dataset (Open, Internal Clinic 1, and Internal Clinic 2) and for both the baseline and BioGuide variants. All models were optimized using the AdamW optimizer. To enhance data variability, we incorporated a data augmentation consisting of random flips, random axial rotations ($\pm 30°$), random intensity shifts, and Gaussian noise injection. Gradient clipping with a maximum norm of 1.0 was applied to stabilize training.

Hyperparameters were selected via a grid search for 200 epochs, while final evaluations were performed for 300 epochs. We evaluated the OneCycle scheduler against a linear warm-up combined with `ReduceLROnPlateau`. The latter was consistently superior. For the BioGuide variant, a learning rate $lr = 10^{-4}$ with a patience equals 20 was optimal. In contrast, the baseline performed best with a higher $lr = 10^{-3}$ and lower patience (10). We compared the hybrid loss function from the baseline study (Dice + Cross-Entropy) against a weighted Dice-Tversky loss. While the hybrid loss was robust, the Dice-Tversky loss with asymmetry weights $\alpha = 0.3$ and $\beta = 0.7$ was selected for all configurations as superior. The auxiliary classification head was optimized by varying the loss weight ($\lambda \in [0.5, 0.1, 0.05, 0.01]$), MLP depth ($L_{MLP} \in [2, 3, 4]$), and hidden dimension ($H_{MLP} \in [64, 128, 266]$). Across all datasets, $\lambda = 0.05$ and $L_{MLP} = 2$ were selected. The hidden dimension H_{MLP} was 256 for the Open dataset and 128 for both internal clinics.

Table 1. Comprehensive segmentation results across Multiple Sclerosis (2D nnU-Net) and various FCD internal/open datasets (Transformer) using BioGuide. Results are reported as mean ± standard deviation across 5-fold cross-validation. Dice, Precision and Detection Rate (Det. Rate) are shown in (%).

Dataset	Set	Method	Dice ↑	Precision ↑	Det. Rate ↑
MS (2D nnU-Net)	V	Baseline (nnU-Net)	67.5 ± 3.5	76.1 ± 0.9	69.4 ± 3.4
		BioGuide (Ours)	$\mathbf{68.6 \pm 2.3}$	$\mathbf{77.6 \pm 0.5}$	$\mathbf{71.5 \pm 2.9}$
FCD Internal 1	V	Baseline Transformer	$32.8_{\pm 6.9}$	$30.2_{\pm 8.3}$	$36.8_{\pm 12.3}$
		Random Noise Vector	$32.9_{\pm 6.6}$	$29.7_{\pm 7.4}$	$32.8_{\pm 13.8}$
		Shuffled Medical Labels	$33.2_{\pm 6.6}$	$31.9_{\pm 8.1}$	$39.8_{\pm 7.5}$
		BioGuide (Location Only)	$33.8_{\pm 6.1}$	$30.5_{\pm 6.5}$	$48.2_{\pm 10.8}$
		BioGuide (Signs Only)	$35.2_{\pm 5.2}$	$32.2_{\pm 5.0}$	$47.3_{\pm 10.6}$
		BioGuide (Full)	$\mathbf{36.1_{\pm 4.6}}$	$\mathbf{32.5_{\pm 5.6}}$	$\mathbf{48.8_{\pm 14.8}}$
	T	Baseline Transformer	$18.5_{\pm 2.3}$	$22.6_{\pm 1.7}$	$51.2_{\pm 2.2}$
		Random Noise Vector	$19.0_{\pm 1.1}$	$23.0_{\pm 0.5}$	$50.5_{\pm 9.6}$
		Shuffled Medical Labels	$19.0_{\pm 1.6}$	$24.4_{\pm 2.5}$	$52.5_{\pm 9.2}$
		BioGuide (Location Only)	$20.3_{\pm 2.2}$	$25.6_{\pm 3.1}$	$45.3_{\pm 5.7}$
		BioGuide (Signs Only)	$20.1_{\pm 2.2}$	$26.3_{\pm 3.4}$	$53.2_{\pm 8.6}$
		BioGuide (Full)	$\mathbf{21.0_{\pm 2.0}}$	$\mathbf{27.7_{\pm 2.4}}$	$\mathbf{58.2_{\pm 7.1}}$
FCD Internal 2	V	Baseline Transformer	$43.0_{\pm 1.8}$	$40.5_{\pm 3.4}$	$70.9_{\pm 13.6}$
		Random Noise Vector	$44.1_{\pm 3.3}$	$43.1_{\pm 5.2}$	$72.1_{\pm 10.7}$
		Shuffled Medical Labels	$44.0_{\pm 5.4}$	$42.6_{\pm 7.6}$	$74.5_{\pm 9.7}$
		BioGuide (Location Only)	$44.4_{\pm 4.7}$	$43.6_{\pm 7.1}$	$74.5_{\pm 9.7}$
		BioGuide (Signs Only)	$\mathbf{45.7_{\pm 4.9}}$	$43.9_{\pm 8.6}$	$73.3_{\pm 9.0}$
		BioGuide (Full)	$\mathbf{45.7_{\pm 4.0}}$	$\mathbf{45.0_{\pm 6.0}}$	$\mathbf{75.8_{\pm 6.0}}$
	T	Baseline Transformer	$37.5_{\pm 1.1}$	$30.9_{\pm 2.7}$	$72.0_{\pm 14.1}$
		Random Noise Vector	$36.7_{\pm 2.2}$	$29.8_{\pm 2.5}$	$64.0_{\pm 8.9}$
		Shuffled Medical Labels	$38.5_{\pm 3.3}$	$31.5_{\pm 3.8}$	$70.0_{\pm 11.5}$
		BioGuide (Location Only)	$37.8_{\pm 2.2}$	$31.4_{\pm 2.7}$	$68.0_{\pm 11.0}$
		BioGuide (Signs Only)	$40.3_{\pm 3.0}$	$33.6_{\pm 2.9}$	$72.0_{\pm 11.0}$
		BioGuide (Full)	$\mathbf{40.4_{\pm 2.3}}$	$\mathbf{34.5_{\pm 2.7}}$	$\mathbf{80.0_{\pm 11.4}}$
FCD Open Dataset	V	Baseline Transformer	$34.0_{\pm 7.5}$	$33.7_{\pm 8.0}$	$54.4_{\pm 5.7}$
		Random Noise Vector	$34.3_{\pm 2.0}$	$34.0_{\pm 4.2}$	$50.7_{\pm 10.2}$
		Shuffled Medical Labels	$36.1_{\pm 5.1}$	$35.7_{\pm 7.0}$	$54.6_{\pm 6.0}$
		BioGuide (Location Only)	$37.2_{\pm 6.3}$	$36.5_{\pm 5.0}$	$55.6_{\pm 6.0}$
		BioGuide (Signs Only)	$38.5_{\pm 4.3}$	$37.2_{\pm 5.4}$	$58.5_{\pm 13.9}$
		BioGuide (Full)	$39.0_{\pm 5.2}$	$38.2_{\pm 6.0}$	$59.7_{\pm 15.9}$

MS Optimization and Framework Integration. For the MS task, we utilized the `nnU-Net` framework to demonstrate the plug-and-play nature of BioGuide. Because `nnU-Net` automatically determines architectural parameters, learning rates, and scheduling, our search was restricted to the BioGuide-specific parameters. The optimal configuration mirrored the most complex FCD setup: $\lambda = 0.05$, a 2-layer MLP, and a hidden dimension of 256. All other training procedures remained fixed to the nnU-Net's self-configured defaults.

3.4 Evaluation Metrics

Performance is evaluated using metrics that capture both spatial overlap and clinical relevance. The **Dice Similarity Coefficient (DSC)** and **Precision**

Table 2. Dice score (%) sensitivity to training set size (70%–100%)

Dataset	Set	Method	70%	80%	90%	100%
Internal 1	V	Baseline	$28.4_{\pm 7.4}$	$29.3_{\pm 8.1}$	$30.9_{\pm 8.2}$	$32.8_{\pm 6.9}$
		BioGuide	$\mathbf{35.1_{\pm 5.8}}$	$\mathbf{32.5_{\pm 8.6}}$	$\mathbf{35.9_{\pm 4.4}}$	$\mathbf{36.1_{\pm 4.6}}$
	T	Baseline	$18.0_{\pm 2.4}$	$18.0_{\pm 1.1}$	$17.4_{\pm 1.3}$	$18.5_{\pm 2.3}$
		BioGuide	$\mathbf{18.2_{\pm 2.1}}$	$\mathbf{18.9_{\pm 1.9}}$	$\mathbf{19.3_{\pm 2.0}}$	$\mathbf{21.0_{\pm 2.0}}$
Internal 2	V	Baseline	$36.8_{\pm 5.6}$	$39.9_{\pm 4.1}$	$40.5_{\pm 3.0}$	$43.0_{\pm 1.8}$
		BioGuide	$\mathbf{39.1_{\pm 8.6}}$	$\mathbf{42.6_{\pm 6.7}}$	$\mathbf{42.7_{\pm 4.3}}$	$\mathbf{45.7_{\pm 4.0}}$
	T	Baseline	$34.0_{\pm 5.7}$	$35.8_{\pm 6.5}$	$36.9_{\pm 2.4}$	$37.5_{\pm 1.1}$
		BioGuide	$\mathbf{35.7_{\pm 3.1}}$	$\mathbf{38.6_{\pm 4.4}}$	$\mathbf{39.1_{\pm 2.3}}$	$\mathbf{40.4_{\pm 2.3}}$
Open Dataset	V	Baseline	$33.1_{\pm 6.7}$	$31.7_{\pm 5.7}$	$31.3_{\pm 6.6}$	$34.0_{\pm 7.5}$
		BioGuide	$\mathbf{36.0_{\pm 6.7}}$	$\mathbf{38.7_{\pm 3.9}}$	$\mathbf{39.2_{\pm 5.7}}$	$\mathbf{39.0_{\pm 5.2}}$

are calculated at the voxel level to quantify segmentation accuracy: $DSC = \frac{2TP}{2TP+FP+FN}$, $Precision = \frac{TP}{TP+FP}$. To assess clinical utility, we employ task-specific Detection Rates: for MS and subjects with multiple lesions, this is the proportion of non-connected clusters successfully detected relative to total ground-truth lesions; for FCD with one lesion, it is the percentage of cases where the single focal lesion is correctly detected.

3.5 Results

MS pathology. The integration of location-based auxiliary labels within the nnU-Net framework yielded improvements across all primary metrics. As shown in Table 1, BioGuide increased the Dice score from 67.5% to 68.6% and Precision from 76.1% to 77.6%. Notably, the guided model also improved the lesion detection count, reaching 71.5% compared to the baseline's 69.4%. These gains are modest but consistent across all folds: MS lesions are high-contrast on FLAIR and the nnU-Net baseline is already near-optimal, whereas the subtle, heterogeneous FCD lesions leave more room for clinical priors to contribute.

FCD pathology. For the FCD, the impact of clinical guidance was evaluated across three datasets. As detailed in Table 1, the **BioGuide (Full)** configuration achieved the highest performance in every test (T) and validation (V) sets. BioGuide consistently achieved superior results on test (T) sets, improving Dice scores from 18.5% to 21.0% on Internal 1 and from 37.5% to 40.4% on Internal 2, with the latter reaching a peak detection rate of 80.0%. Knowledge-blind ablations performed significantly worse than the full mode, confirming that the improvements stem from genuine biomedical knowledge rather than a simple regularization effect from the additional MLP head. Ablation results indicate that while both **Location-Only** and **Signs-Only** guidance provide benefits, their combination in the Full model consistently maximizes the Percent of Detection, reaching 80.0% in the Internal 2 Test set compared to 72.0% for the baseline.

An additional strength of the BioGuide framework is its performance in limited-sample settings. Table 2 illustrates that BioGuide models trained on

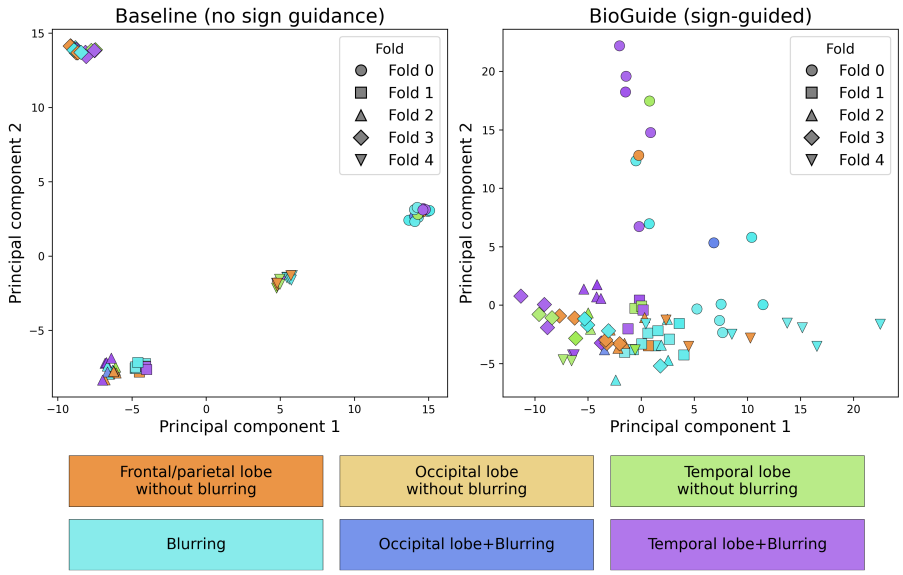

Fig. 3. PCA visualization of features extracted from the encoder bottleneck for both the Baseline (left) and BioGuide (right) using the FCD Internal 2 dataset. BioGuide forces a distinct clustering of the latent space based on radiological descriptors, such as the presence of blurring and anatomical location.

only 80% of the data—with randomly selected training subjects—often outperform baseline models trained on the full 100%. For the Open Dataset, BioGuide achieves a Dice score of 36.0% using only 70% of the training data, surpassing the baseline's performance even when trained on the full dataset (34.0%). This suggests that embedding domain knowledge provides a powerful inductive bias, significantly reducing the annotation burden.

To understand how the auxiliary classification task influences feature learning, we visualized the encoder's bottleneck features using Principal Component Analysis (PCA). As shown in Fig. 3, the baseline model's features remain clustered by folds and overlapping across different clinical phenotypes (different colors), suggesting a lack of semantic differentiation. In contrast, the **BioGuide** bottleneck features demonstrate a clear restructuring of the latent space.

4 Discussion and Conclusion

The BioGuide framework demonstrates that integrating structured clinical descriptors as auxiliary targets effectively regularizes feature extraction in medical segmentation tasks. By enforcing a relationship between latent features and radiological signs—such as lesion location for MS or specific textural markers for FCD—the model develops a more robust inductive bias than pixel-wise supervision alone. Results across four diverse datasets confirm that this guidance consis-

tently improves spatial overlap and clinical detection rates. Moreover, BioGuide demonstrates high data efficiency, outperforming baselines with 30% less data. Furthermore, the knowledge-blind ablation studies using random noise and shuffled labels confirm that these improvements are driven by the semantic alignment with medical truth rather than simple architectural changes.

In conclusion, BioGuide provides a versatile framework for integrating expert knowledge into deep learning architectures. Although validated on brain MRI, its modality-agnostic design extends readily to other clinical domains with structured radiological reports.

While BioGuide performs well, its bottleneck-only design may limit high-resolution feature refinement, motivating future work on multi-scale architectures with auxiliary heads distributed across the decoder. Additionally, the reliance on binary labels and subjective radiological descriptors—which can introduce inter-rater noise—motivates exploring probabilistic soft labels to better capture the continuous clinical spectrum. Future work will also investigate cross-pathology transfer learning to assess whether clinical textures learned in one domain, such as cortical blurring, can improve segmentation in other pathologies without extensive retraining.

Acknowledgments The work was supported by the Russian Science Foundation grant № 25-71-10088.

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

References

1. Amyar, A., Modzelewski, R., Vera, P., Morard, V., Ruan, S.: Multi-task multi-scale learning for outcome prediction in 3d pet images. Computers in biology and medicine **151**(Pt A), 106208 (2022). https://doi.org/10.1016/j.compbiomed.2022.106208
2. Bernasconi, A., Cendes, F., Theodore, W., Gill, R.S., Koepp, M.J., Hogan, R.E., Jackson, G., Federico, P., Labate, A., Vaudano, A.E., et al.: Recommendations for the use of structural MRI in the care of patients with epilepsy: a consensus report from the ILAE neuroimaging task force. Epilepsia **60**, 1054–1068 (2019)
3. Friedrich, P., Frisch, Y., Cattin, P.C.: Deep generative models for 3D medical image synthesis. arXiv preprint arXiv:2410.17664 (2024). https://doi.org/10.48550/ARXIV.2410.17664
4. Haider, L., et al.: The topography of demyelination and neurodegeneration in the multiple sclerosis brain. Brain **139**, 807–815 (2016)
5. Hwang, S., Kim, H.E.: Self-transfer learning for weakly supervised lesion localization. In: International conference on medical image computing and computer-assisted intervention. pp. 239–246. Springer (2016)
6. Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.P., Heiland, S., Wick, W., et al.: Automated brain extraction of multisequence MRI using artificial neural networks. Human brain mapping **40**(17), 4952–4964 (2019)

7. 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)
8. Kamiri, J., Wambugu, G.M., Oirere, A.M.: Multi-task deep learning in medical image processing: a systematic review. International Journal of Computing Sciences Research (2025)
9. Peng, W., Adeli, E., Bosschieter, T., Park, S.H., Zhao, Q., Pohl, K.M.: Generating realistic brain MRIs via a conditional diffusion probabilistic model. In: International conference on medical image computing and computer-assisted intervention. pp. 14–24. Springer (2023)
10. Schuch, F., Walger, L., Schmitz, M., David, B., Bauer, T., Harms, A., Fischbach, L., Schulte, F., Schidlowski, M., Reiter, J., et al.: An open presurgery MRI dataset of people with epilepsy and focal cortical dysplasia type II. Scientific Data **10**(1), 475 (2023)
11. Sun, Y., Liu, M., Lian, C.: Aepl: Automated and editable prompt learning for brain tumor segmentation. In: 2025 IEEE 22nd international symposium on biomedical imaging (ISBI). pp. 1–4. IEEE (2025)
12. Tan, Z., Madzin, H., Norafida, B., ChongShuang, Y., Sun, W., Nie, T., Cai, F.: Deeppulmotb: A benchmark dataset for multi-task learning of tuberculosis lesions in lung computerized tomography (ct). Heliyon **10**(4) (2024)
13. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage **31**(3), 1116–1128 (2006)
14. Zhang, X., Zhang, Y., Wang, C., Li, L., Zhu, F., Sun, Y., Mo, T., Hu, Q., Xu, J., Cao, D.: Focal cortical dysplasia lesion segmentation using multiscale transformer. Insights into Imaging **15**(1), 222 (2024)
15. Zhu, L., Codella, N., Chen, D., Jin, Z., Yuan, L., Yu, L.: Generative enhancement for 3D medical images. arXiv preprint arXiv:2403.12852 (2024). https://doi.org/10.48550/ARXIV.2403.12852