DDecomp: Structured Uncertainty Decomposition Using Masked Autoencoders for Anomaly Detection in Diffusion MRI

ZhiJia Fu[1], Wei Zhang[1], RuiXi Zheng[1], YiJie Li[1], GuanLin Huang[1], Lipeng Ning[2], Lauren J. O'Donnell[2], Ofer Pasternak[2], and Fan Zhang[1⋆]

[1] University of Electronic Science and Technology of China, China
zhangfanmark@gmail.com
[2] Harvard Medical School, United States

Abstract. Diffusion MRI (dMRI) is an advanced imaging technique that can characterize tissue microstructure and map white matter connections *in vivo*. Anomaly detection of brain lesions, such as tumors, white matter hyperintensities, and intracerebral hemorrhage, is a critical task in dMRI, supporting neurological disease diagnosis and surgical planning. Unsupervised anomaly detection leveraging uncertainty models offers an effective approach for distinguishing diffusion measures of abnormal tissues from those of normal structures. However, this task is complicated by acquisition noise and cross-protocol variability, which can obscure subtle pathological signals. In this study, we propose *DDecomp*, a masked-autoencoder-based framework trained on healthy data that decomposes predictive uncertainty and leverages the epistemic component as an out-of-distribution cue for unsupervised anomaly detection. The key idea is to disentangle uncertainty arising from inherent data noise from uncertainty indicative of model mismatch to pathology, thereby generating an epistemic map that is specific to the anomaly. Once trained, the model generalizes to detect multiple lesion types within a unified framework. In the experiments, we evaluate our method on three diverse clinical cohorts (tumors, white matter hyperintensities, and intracerebral hemorrhage) acquired with different scanners and protocols. Across all cohorts, our approach achieves robust lesion detection and consistently outperforms state-of-the-art baselines.

Keywords: Diffusion MRI · Uncertainty Decomposition · Unsupervised Anomaly Detection · Deep Learning.

1 Introduction

Diffusion MRI (dMRI) is an advanced imaging technique that can characterize tissue microstructure [1] and map white matter connections *in vivo* [2]. dMRI has been widely used to study and monitor many neurological disorders [3]. For example, white matter hyperintensities (WMH), which are common in aging and

⋆ Corresponding author.

vascular diseases; glioma, a type of brain tumor; and intracerebral hemorrhage (ICH), which results from bleeding in the brain. Reliable lesion localization in dMRI-derived scalar maps such as fractional anisotropy (FA) and mean diffusivity (MD) remains difficult due to the following factors [4]. dMRI is sensitive to acquisition noise, scanner hardware, and protocol differences [4, 10]. dMRI also has lower spatial resolution than structural MRI [5, 16], and voxel level lesion labels are costly and often ambiguous [6]. Furthermore, lesion appearance can vary across etiologies, reflecting various types of imaging signals.

Unsupervised anomaly detection (UAD) is an effective approach for distinguishing diffusion measures of abnormal tissues from those of normal structures [6]. It avoids dense lesion labels and can leverage large unlabeled datasets. Specifically, masked autoencoder (MAE) is a widely used approach to enable UAD to learn latent representations by reconstructing masked patches from visible context [7] and is often built on Vision Transformers (ViTs) [8]. MAE-based UAD has been explored in medical imaging settings for lesion detection [9], which provides a promising strategy for dMRI-based anomaly detection. We apply MAE to 3D mean diffusivity (MD) volumes by masking 3D patches and reconstructing the missing regions to produce voxel-wise anomaly maps. However, high reconstruction error can reflect not only the pathology but also acquisition noise or protocol differences [10]. As a result, a high score is not necessarily specific to the abnormal lesion regions.

Uncertainty quantification is another approach for UAD [12]. It involves two types of uncertainty: aleatoric and epistemic. Aleatoric uncertainty captures data noise and protocol variability, while epistemic uncertainty reflects the model's limited knowledge due to finite data or capacity. Bayesian deep learning distinguishes aleatoric and epistemic uncertainty [12]. Monte Carlo (MC) dropout is a practical approximation [13, 14] that samples random subnetworks at test time to estimate predictive uncertainty [13].

In dMRI-related work, the recently proposed DDEvENet [11] has attempted to leverage uncertainty quantification to assist in the identification of abnormal lesions while performing anatomical brain parcellation. However, it does not explicitly separate these two types of uncertainty [11, 12], which can result in uncertainty estimates influenced by both noise and cross-protocol variability [10]. In dMRI, observation noise and cross-protocol shift can be substantial [4, 10], so MC dropout uncertainty can become a mixed signal and may be unreliable for lesion localization. MAE reconstruction also introduces randomness through the mask pattern [7], and lesion scores based on a single mask can be unstable. To our knowledge, uncertainty decomposition has not been explored for dMRI uncertainty maps.

In this study, we propose *DDecomp*, a masked-autoencoder-based framework that makes uncertainty a pathology-sensitive cue for unsupervised anomaly detection in dMRI. The central novelty is to explicitly disentangle uncertainty driven by acquisition noise and cross-protocol variability (aleatoric) from uncertainty caused by model mismatch to out-of-distribution tissue (epistemic), and to leverage the epistemic component for lesion localization. To further address

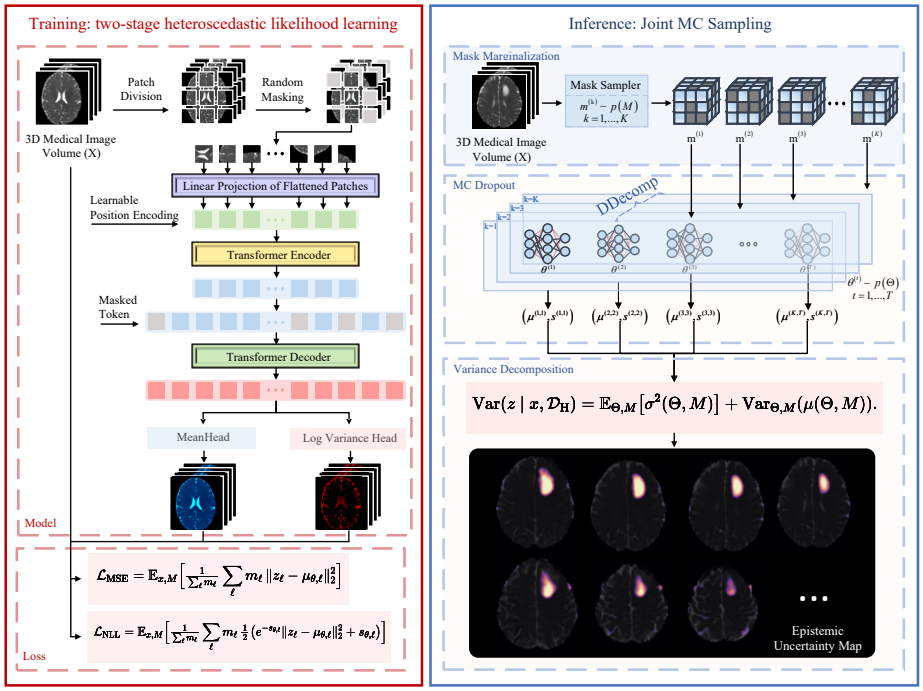

Fig. 1. DDecomp: (left) two-stage training; (right) joint MC inference over dropout and masks.

the instability introduced by stochastic masking in MAEs, we treat the mask as a nuisance variable and marginalize it at inference, yielding a more stable posterior predictive estimate. Built on a ViT-MAE backbone [7, 8] with heteroscedastic modeling, DDecomp produces a calibrated voxel-wise epistemic map via uncertainty decomposition, improving lesion specificity under substantial noise and protocol shift. We additionally introduce an uncertainty signal-to-noise metric to quantify salience in pathological regions, and demonstrate that a single trained model generalizes across three heterogeneous clinical cohorts (glioma, WMH, and ICH) acquired on different scanners and protocols, consistently outperforming state-of-the-art unsupervised baselines.

2 Method

2.1 Preliminaries

Let $x \in \mathbb{R}^{H\times W\times D}$ denote a 3D dMRI-derived scalar volume. We train on a healthy-only dataset $\mathcal{D}_{\mathrm{H}}$, the Human Connectome Project (HCP) [15, 16] and test on pathological cohorts $\mathcal{D}_{\mathrm{P}}$ (e.g., glioma/hemorrhage/WMH), with $\mathcal{D}_{\mathrm{H}} \cap \mathcal{D}_{\mathrm{P}} = \varnothing$. We partition x into non-overlapping cubic patches of size $p \times p \times p$

(assuming H, W, D are divisible by p). Let patchify$(\cdot)$ map x to a sequence of patch tokens:

$$z = \text{patchify}(x) \in \mathbb{R}^{L \times P}, \qquad L = \frac{HWD}{p^3}, \qquad P = p^3 \tag{1}$$

Here $z_\ell \in \mathbb{R}^P$ denotes the flattened voxel vector of patch ℓ (single-channel scalar volume). Let unpatchify$(\cdot)$ denote the inverse mapping back to voxel space. A binary mask $M \in \{0,1\}^L$ indicates visible and masked patch indices:

$$\mathcal{V}(M) = \{\ell \mid M_\ell = 1\}, \qquad \mathcal{M}(M) = \{\ell \mid M_\ell = 0\} \tag{2}$$

Given M, the MAE encoder processes only visible tokens $z_\mathcal{V} = \{z_\ell\}_{\ell \in \mathcal{V}(M)}$, while the decoder receives (i) encoded visible tokens and (ii) learned mask tokens inserted at masked positions prior to reconstruction. We model the conditional reconstruction distribution over masked tokens as

$$p_\theta(z_\mathcal{M} \mid z_\mathcal{V}, M) \tag{3}$$

where θ denotes learned network parameters.

2.2 Training: Two-stage heteroscedastic likelihood learning

To explicitly model observation noise and cross-protocol variability, we equip the decoder with two heads: a mean head $\mu_\theta(\cdot)$ and a log-variance head $s_\theta(\cdot)$:

$$\mu_\theta(z_\mathcal{V}, M) \in \mathbb{R}^{L \times P}, \qquad s_\theta(z_\mathcal{V}, M) \in \mathbb{R}^{L \times P}, \qquad \sigma_\theta^2 = \exp\big(s_\theta\big) \tag{4}$$

We assume a diagonal Gaussian likelihood on the masked tokens, factorized across masked patches and within-patch voxels:

$$p_\theta(z_\mathcal{M} \mid z_\mathcal{V}, M) = \prod_{\ell \in \mathcal{M}(M)} \prod_{j=1}^{P} \mathcal{N}\Big(z_{\ell j};\, \mu_{\theta,\ell j}(z_\mathcal{V}, M),\, \sigma^2_{\theta,\ell j}(z_\mathcal{V}, M)\Big) \tag{5}$$

Stage I (mean reconstructor) We learn a strong conditional mean reconstructor via masked MSE (standard MAE objective [7]):

$$\theta_{\text{MSE}} = \arg\min_\theta \; \mathbb{E}_{x \sim \mathcal{D}_\text{H}} \, \mathbb{E}_{M \sim p_\rho(M)} \big[\mathcal{L}_{\text{MSE}}(\theta; x, M)\big], \tag{6}$$

$$\mathcal{L}_{\text{MSE}}(\theta; x, M) = \frac{1}{|\mathcal{M}(M)|\, P} \sum_{\ell \in \mathcal{M}(M)} \|z_\ell - \mu_{\theta,\ell}(z_\mathcal{V}, M)\|_2^2 \tag{7}$$

where $p_\rho(M)$ is the masking distribution with mask ratio ρ.

Stage II (heteroscedastic likelihood) Initialized from θ_{MSE}, we fine-tune both heads under the heteroscedastic Gaussian NLL [12]:

$$\theta = \arg\min_{\theta} \; \mathbb{E}_{x\sim\mathcal{D}_{\mathrm{H}}} \, \mathbb{E}_{M\sim p_\rho(M)} \big[\mathcal{L}_{\mathrm{NLL}}(\theta; x, M)\big], \tag{8}$$

$$\mathcal{L}_{\mathrm{NLL}}(\theta; x, M) = \frac{1}{|\mathcal{M}(M)|\, P} \sum_{\ell\in\mathcal{M}(M)} \sum_{j=1}^{P} \frac{1}{2}\Big(\exp(-s_{\theta,\ell j})\big(z_{\ell j} - \mu_{\theta,\ell j}\big)^2 + s_{\theta,\ell j}\Big) \tag{9}$$

where $\mu_{\theta,\ell j}$ and $s_{\theta,\ell j}$ are shorthand for $\mu_{\theta,\ell j}(z_{\mathcal{V}}, M)$ and $s_{\theta,\ell j}(z_{\mathcal{V}}, M)$, respectively, and constants independent of θ are omitted.

2.3 Inference: Joint MC sampling

To estimate the total uncertainty of the model, we treat network parameters as a random variable Θ with posterior $p(\Theta \mid \mathcal{D}_{\mathrm{H}})$. Exact Bayesian inference is intractable, so we use MC Dropout [13] as a variational approximation: enabling dropout at test time induces a distribution $q(\Theta)$ around the trained parameters (dropout regularization follows the standard formulation [13]). Concretely, we obtain parameter samples $\Theta^{(t)} \sim q(\Theta)$ by performing stochastic forward passes with dropout enabled.

In addition, because MAE reconstruction is conditioned on a stochastic mask M, we treat M as a nuisance condition and marginalize it via MC sampling. The mask-marginalized posterior predictive over masked tokens is:

$$\begin{aligned} p(z_{\mathcal{M}} \mid x, \mathcal{D}_{\mathrm{H}}) &= \int p_\rho(M) \left(\int p(z_{\mathcal{M}} \mid z_{\mathcal{V}}, M, \Theta)\, p(\Theta \mid \mathcal{D}_{\mathrm{H}})\, d\Theta \right) dM \\ &\approx \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} p\Big(z_{\mathcal{M}} \mid z_{\mathcal{V}}, M^{(k)}, \Theta^{(t)}\Big) \end{aligned} \tag{10}$$

where $M^{(k)} \sim p_\rho(M)$ and $\Theta^{(t)} \sim q(\Theta)$ are sampled by random masks and dropout-enabled forward passes, respectively.

Each stochastic forward pass yields a Gaussian component with mean and variance maps:

$$\begin{aligned} p\Big(z_{\mathcal{M}} \mid z_{\mathcal{V}}, M^{(k)}, \Theta^{(t)}\Big) &= \mathcal{N}\Big(z_{\mathcal{M}};\, \mu^{(t,k)},\, \mathrm{diag}(\sigma^{2(t,k)})\Big)\,, \\ \mu^{(t,k)} &= \mu_{\Theta^{(t)}}(z_{\mathcal{V}}, M^{(k)}), \\ \sigma^{2(t,k)} &= \exp\big(s_{\Theta^{(t)}}(z_{\mathcal{V}}, M^{(k)})\big) \end{aligned} \tag{11}$$

2.4 Uncertainty decomposition

Applying the law of total variance to the joint randomness of (Θ, M) yields:

$$\begin{aligned} \mathrm{Var}(z \mid x, \mathcal{D}_{\mathrm{H}}) &= \mathbb{E}_{\Theta,M}\big[\mathrm{Var}(z \mid \Theta, M)\big] + \mathrm{Var}_{\Theta,M}\big(\mathbb{E}[z \mid \Theta, M]\big) \\ &= \mathbb{E}_{\Theta,M}\big[\sigma^2(\Theta, M)\big] + \mathrm{Var}_{\Theta,M}\big(\mu(\Theta, M)\big) \end{aligned} \tag{12}$$

We estimate these terms from KT samples. The empirical predictive mean is:

$$\widehat{\mu} = \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} \mu^{(t,k)} \tag{13}$$

The aleatoric estimator (heteroscedastic noise term) is:

$$\widehat{U}_{\text{alea}} = \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} \sigma^{2(t,k)} \tag{14}$$

The epistemic estimator (variance of predictive means across subnetworks and mask patterns) is:

$$\widehat{U}_{\text{epi}} = \left(\frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} \left(\mu^{(t,k)}\right)^2 \right) - \widehat{\mu}^{\,2} \tag{15}$$

A moment-matching estimate of the total predictive variance is:

$$\widehat{U}_{\text{total}} = \widehat{U}_{\text{alea}} + \widehat{U}_{\text{epi}} \tag{16}$$

We use $\widehat{U}_{\text{epi}}$ as a noise-corrected epistemic map for lesion localization: it isolates variability of the predicted mean under changes in (Θ, M), while $\widehat{U}_{\text{alea}}$ captures heteroscedastic observation noise learned from healthy data. Thus, epistemic uncertainty reflects model mismatch to the healthy training distribution, while aleatoric uncertainty captures acquisition/protocol variability. Severe preprocessing failures may still cause epistemic responses and should be excluded by QC(Quality Control).

2.5 Why epistemic uncertainty detects anomaly?

Healthy samples concentrate near a low-dimensional manifold $\mathcal{M}_{\text{H}}$. Training on $\mathcal{D}_{\text{H}}$ makes the model confident around $\mathcal{M}_{\text{H}}$, while inputs that deviate from it (e.g., lesions) induce larger variability of the predictive mean across stochastic subnetworks and mask patterns. Under a local linearization $\mu_{\theta,M}(u) \approx \bar{\mu}_M(u) + \bar{J}_M(u)(\theta - \bar{\theta})$, the epistemic component approximately follows

$$\mathrm{Var}_\theta\big(\mu_{\theta,M}(u)\big) \approx \bar{J}_M(u)\, \Sigma_\theta\, \bar{J}_M(u)^\top, \tag{17}$$

so off-manifold regions naturally yield larger $\bar{J}_M(u)$ and hence higher epistemic contrast. Meanwhile, acquisition/protocol variability is captured by the heteroscedastic head as an aleatoric term σ^2; separating it prevents nuisance variance from diluting lesion salience.

3 Experiments and Results

3.1 Experiments

Data. We trained on 960 healthy HCP subjects [15, 16] and tested on three clinical cohorts: 11 WMH patients from the Second Affiliated Hospital, Zhejiang University School of Medicine (62.6 ± 19.8 years; 2F/9M; 3T GE MR750; TE/TR = 80.8/8000 ms; 30 directions); 22 glioma patients from the First Affiliated Hospital of Sun Yat-sen University (48.1 ± 23.1 years; 9F/13M; 3T Siemens Prisma; TE/TR = 79.0/22000 ms; 1 $b = 0$, 64 directions); and 20 ICH patients (GE Discovery MR750; TE/TR = 80.8/8750 ms; 1 $b = 0$, 30 directions). All clinical dMRI scans used $b = 1000$ s/mm^2. No HCP subjects were manually excluded for pathologies, so rare incidental abnormalities may remain.

For each dMRI scan, we computed a diffusion tensor-derived MD map as network input using SlicerDMRI [17–19]. The baseline image was rigidly registered to an MNI template using ANTs, and the resulting transform was applied to MD [20, 21]. The MD map was upsampled to 1 mm isotropic voxels with a matrix size of $192 \times 256 \times 256$.

Implementation details. We trained our model on a single NVIDIA RTX 4090 GPU. In stage one, we trained for 4000 epochs with batch size 8. We used the Adam optimizer with a learning rate 0.0001 and a mask ratio 0.75 [23]. In stage two, we trained for 200 epochs with batch size 8. Inference used $K = 100$, $T = 10$, and took about 2 min per subject on RTX 4090. The implementation was based on PyTorch [22].Code and weights will be released at `https://github.com/WHOOOOOS/DDecomp`.

Baselines and evaluation metrics. Prior dMRI-based studies often focus on unsupervised lesion detection using reconstruction error, or on uncertainty outputs from uncertainty quantification models. We refer to both as *anomaly scores*. We compared our method with the uncertainty quantification model DDEvENet [11] and an MAE-based UAD baseline, denoted as MAE-UAD.

We threshold each anomaly map into a binary lesion mask and compute F_β against ground truth, with $\beta = 3$ to emphasize recall; AUROC is also reported [24, 25]. Anomaly scores can be affected by acquisition noise, scanner hardware, and protocol differences, and high responses can also appear near some brain boundaries. AUROC is less sensitive to this behavior.

To reflect the strength of effective signal relative to background response, we designed a SNR metric for comparing the salience of lesions in anomaly maps across DDecomp, DDEvENet, and MAE-UAD. We define SNR as the sum of anomaly values within the lesion region divided by the sum of anomaly values within the non-lesion region:

$$\mathrm{SNR} = \frac{\sum_{v \in \Omega_{\mathrm{les}}} a(v)}{\sum_{v \in \Omega_{\mathrm{non}}} a(v)}. \tag{18}$$

Table 1. Comparison of DDEvENet, MAE-UAD, and DDecomp on glioma, WMH, and ICH. AUROC, F_3, and SNR are reported. For fairness, each model uses its selected anomaly threshold when computing F_3.

Models	glioma			WMH			ICH		
	AUROC	F_3	SNR	AUROC	F_3	SNR	AUROC	F_3	SNR
DDEvENet	0.86 ±0.07	0.36 ±0.09	0.04 ±0.02	0.89 ±0.04	0.21 ±0.06	0.02 ±0.01	0.87 ±0.10	0.36 ±0.12	0.04 ±0.02
MAE-UAD	**0.95 ±0.04**	0.66 ±0.16	0.33 ±0.03	0.92 ±0.04	0.40 ±0.11	0.06 ±0.03	0.98 ±0.02	0.67 ±0.22	0.23 ±0.05
DDecomp	0.95 ±0.05	**0.75 ±0.12**	**1.83 ±0.16**	**0.95 ±0.04**	**0.59 ±0.15**	**0.27 ±0.07**	**0.99 ±0.01**	**0.80 ±0.14**	**0.84 ±0.18**

Here $a(v)$ is the anomaly value at voxel v, Ω_{les} denotes lesion voxels, and Ω_{non} denotes non-lesion brain voxels.

3.2 Results

Table 1 reports detection performance of DDEvENet, MAE-UAD, and DDecomp on the glioma, WMH, and ICH datasets. Our method remains better than the baselines on AUROC, F_3 and also achieves substantially higher SNR than DDEvENet and MAE-UAD. This pattern is consistent with more effective separation of uncertainty components and a cleaner epistemic signal for unsupervised lesion detection.

Figure 2 shows axial 2D slices of anomaly maps for the three datasets. Visually, DDecomp is better than MAE-UAD and DDEvENet. The epistemic uncertainty map from DDecomp highlights the lesion region more directly, and it is less affected by responses related to acquisition noise, scanner hardware, and protocol differences. DDEvENet performs uncertainty estimation based on a segmentation model, so it is sensitive to spatial shifts near tissue boundaries. This observation supports MAE as a suitable backbone for uncertainty quantification and decomposition in this setting.

4 Discussion and Conclusion

We propose DDecomp, a healthy-data-trained MAE framework for dMRI that decomposes predictive uncertainty by jointly marginalizing dropout and mask sampling. By separating aleatoric noise from epistemic uncertainty, DDecomp yields a calibrated epistemic map that is more pathology-specific. Experiments on glioma, WMH, and ICH show consistent improvements over DDEvENet and MAE-UAD in AUROC, F_3, and SNR, with qualitative results indicating clearer lesion salience and less sensitivity to noise and protocol differences. DDecomp is a general uncertainty-decomposition framework, although this work focuses on dMRI where acquisition variability and microstructural pathology make aleatoric/epistemic separation important. Limitations include reliance on

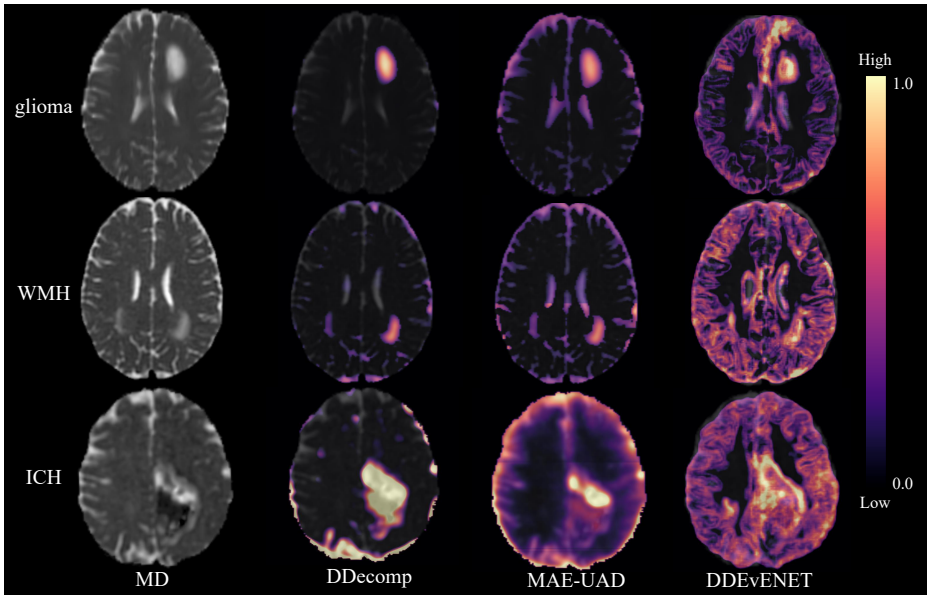

Fig. 2. Axial 2D slices of anomaly maps for the three datasets. Row 1 corresponds to glioma, row 2 corresponds to WMH, and row 3 corresponds to ICH. Column 1 shows the original MD map. Column 2 shows DDecomp. Column 3 shows MAE-UAD. Column 4 shows DDEvENet.

quality-controlled preprocessing, possible epistemic responses to severe artifacts, rare incidental abnormalities in HCP, and limited baseline comparisons. Future work will test other modalities and broader UAD methods.

Acknowledgments This work is in part supported by the National Key R&D Program of China (No. 2023YFE0118600), the National Natural Science Foundation of China (No. 62371107), and Science and Technology Department of Sichuan Province (No. 2026YFHZ0045).

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

References

1. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI. *Journal of Magnetic Resonance* **213**(2), 560–570 (2011). https://doi.org/10.1016/j.jmr.2011.09.022
2. Zhang, F., Daducci, A., He, Y., Schiavi, S., Seguin, C., Smith, R.E., Yeh, C.H., Zhao, T., O'Donnell, L.J.: Quantitative mapping of the brain's structural connectivity using diffusion MRI tractography: A review. *NeuroImage* **249**, 118870 (2022). https://doi.org/10.1016/j.neuroimage.2021.118870

3. Liao, Y., Coelho, S., Chen, J., Ades-Aron, B., Pang, M., Osorio, R., Shepherd, T., Lui, Y.W., Novikov, D.S., Fieremans, E.: Mapping tissue microstructure of brain white matter in vivo in health and disease using diffusion MRI. *Imaging Neuroscience* **2**, 1–17 (2024). https://doi.org/10.1162/imag_a_00102. arXiv:2307.16386
4. Jones, D.K., Cercignani, M.: Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR in Biomedicine* **23**(7), 803–820 (2010). https://doi.org/10.1002/nbm.1543
5. Malinsky, M., Peter, R., Hodneland, E., et al.: Registration of FA and T1-weighted MRI data of healthy human brain based on template matching and normalized cross-correlation. *Journal of Digital Imaging* **26**, 774–785 (2013). https://doi.org/10.1007/s10278-012-9561-8
6. Baur, C., Denner, S., Wiestler, B., Navab, N., Albarqouni, S.: Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study. *Medical Image Analysis* **69**, 101952 (2021). https://doi.org/10.1016/j.media.2020.101952
7. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: *Proc. CVPR*, pp. 16000–16009 (2022). arXiv:2111.06377
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al.: An image is worth 16×16 words: Transformers for image recognition at scale. In: *ICLR* (2021). arXiv:2010.11929
9. Georgescu, M.-I.: Masked autoencoders for unsupervised anomaly detection in medical images. arXiv:2307.07534 (2023)
10. Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., et al.: Multi-site harmonization of diffusion MRI data in a registration framework. *Brain Imaging and Behavior* **12**(1), 284–295 (2018). https://doi.org/10.1007/s11682-016-9670-y
11. Li, C., Yang, D., Yao, S., Wang, S., Wu, Y., Zhang, L., et al.: DDEvENet: Evidence-based ensemble learning for uncertainty-aware brain parcellation using diffusion MRI. arXiv:2409.07020 (2024)
12. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: *Advances in Neural Information Processing Systems* (NeurIPS) (2017). arXiv:1703.04977
13. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: *Proc. ICML* (2016). arXiv:1506.02142
14. Gal, Y.: *Uncertainty in Deep Learning*. PhD thesis, University of Cambridge (2016)
15. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K.: The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80**, 62–79 (2013). https://doi.org/10.1016/j.neuroimage.2013.05.041
16. Sotiropoulos, S.N., Jbabdi, S., Xu, J., et al.: Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage* **80**, 125–143 (2013). https://doi.org/10.1016/j.neuroimage.2013.05.057
17. Norton, I., Essayed, W.I., Zhang, F., Pujol, S., Yarmarkovich, A., Golby, A.J., Kindlmann, G., Wassermann, D., Estepar, R.S.J., Rathi, Y., Pieper, S., Kikinis, R., Johnson, H.J., Westin, C.-F., O'Donnell, L.J.: SlicerDMRI: Open source diffusion MRI software for brain cancer research. *Cancer Research* **77**(21), e101–e103 (2017). https://doi.org/10.1158/0008-5472.CAN-17-0332
18. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., et al.: 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging* **30**(9), 1323–1341 (2012). https://doi.org/10.1016/j.mri.2012.05.001

19. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. *Biophysical Journal* **66**(1), 259–267 (1994). https://doi.org/10.1016/S0006-3495(94)80775-1
20. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* **54**(3), 2033–2044 (2011). https://doi.org/10.1016/j.neuroimage.2010.09.025
21. Fonov, V.S., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., et al.: Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage* **54**(1), 313–327 (2011). https://doi.org/10.1016/j.neuroimage.2010.07.033
22. Paszke, A., Gross, S., Massa, F., et al.: PyTorch: An imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems*, vol. 32, pp. 8024–8035 (2019). arXiv:1912.01703
23. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: *ICLR* (2015). arXiv:1412.6980
24. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves. In: *Proc. ICML*, pp. 233–240 (2006). https://doi.org/10.1145/1143844.1143874
25. Fawcett, T.: An introduction to ROC analysis. *Pattern Recognition Letters* **27**(8), 861–874 (2006). https://doi.org/10.1016/j.patrec.2005.10.010