Federated Disentangled Representation Learning for Unsupervised Brain Anomaly Detection

Cosmin Bercea

Helmholtz Zentrum München

Benedikt Wiestler

Technical University Munich

Daniel Rueckert

Technical University of Munich

Shadi Albarqouni (✉ shadi.albarqouni@helmholtz-muenchen.de)

Helmholtz Zentrum München <https://orcid.org/0000-0003-2157-2211>

Article

Keywords: FedDis, abnormal brain MRIs, unsupervised deep convolutional neural network

Posted Date: August 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-722389/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Machine Intelligence on August 25th, 2022. See the published version at <https://doi.org/10.1038/s42256-022-00515-2>.

Federated Disentangled Representation Learning for Unsupervised Brain Anomaly Detection

Cosmin I. Bercea^{1,2}, Benedikt Wiestler⁴, Daniel Rueckert^{2,3,4}, and Shadi Albarqouni^{1,2,*}

¹Helmholtz AI, Helmholtz Zentrum München, Neuherberg, Germany

²Technical University of Munich, Munich, Germany

³Imperial College London, United Kingdom

⁴Klinikum Rechts der Isar, Munich, Germany

*Corresponding author: shadi.albarqouni@helmholtz-muenchen.de

ABSTRACT

Recent advances in Deep Learning (DL) and the increased use of brain MRI have provided a great opportunity and interest in automated anomaly segmentation to support human interpretation and improve clinical workflow. However, medical imaging must be curated by trained clinicians, which is time-consuming and expensive. Further, data is often scattered across multiple institutions, with privacy regulations limiting its access. Here, we present FedDis (Federated Disentangled representation learning for unsupervised brain pathology segmentation) to collaboratively train an unsupervised deep convolutional neural network on 1532 healthy MR scans from four different institutions, and evaluate its performance in identifying abnormal brain MRIs including multiple sclerosis (MS) lesions, low-grade tumors (LGG), and high-grade tumors/glioblastoma (HGG/GB) on a total of ≈ 500 scans from 5 different institutions and datasets. FedDis mitigates the statistical heterogeneity given by different scanners by disentangling the parameter space into global, i.e., shape and local, i.e., appearance. We only share the former with the federated clients to leverage common anatomical structure while keeping client-specific contrast information private. We have shown that our collaborative approach, FedDis, improves anomaly segmentation results by 99.74% for MS and 40.45% for tumors over locally trained models without the need for annotations or sharing private local data. We found out that FedDis is especially beneficial for clients that share both healthy and anomaly data coming from the same institute, improving their local anomaly detection performance by up to 227% for MS lesions and 77% for brain tumors.

1 Introduction

Brain magnetic resonance imaging (MRI) is one of the most commonly used tests in neurology and neurosurgery. Due to the high contrast of the soft tissues, different MR sequences, e.g., Fluid Attenuated Inversion Recovery (FLAIR) images are very sensitive to pathology and are useful in detecting several abnormalities such as tumors, inflammation, multiple sclerosis, or acute infarctions.

Multiple sclerosis is an immune-mediated inflammatory disease that destroys myelin and axons, resulting in significant physical disability. MRI is the most commonly used modality for the diagnosis of MS lesions and assessment of disease progression⁹. The most used method for detecting MS lesions in clinical practice is to threshold the hyper-intense regions in an MRI scan, e.g., FLAIR. However, different residual and intensity artifacts make it challenging to compute the MS lesion volume. Therefore, the development of robust and automated MS lesion detection methods can reduce the burden on radiologists and improve diagnostic performance.

Glioma is an aggressive type of brain cancer classified by its histopathological appearances into low-grade glioma (grades I and II) and high-grade glioma (grades III and IV). Despite improvement in the diagnosis and treatment, brain tumors are still associated with significant morbidity and a poor overall prognosis. Automatic assessments of the tumor boundary and volume are beneficial for treatment planning and monitoring treatment response¹⁰.

Recent technological advances have resulted in much faster and higher resolution MR imaging, which has led to an overburdening of the radiologist in the interpretation and triage of acute findings. One solution to address these challenges is the use of artificial intelligence for the automated segmentation of abnormalities. These methods could provide rapid interpretation of brain MR images, improving clinical workflow and thus reducing the burden on the radiologist.

Even though machine learning methods have been successfully applied in various medical applications, their performance is limited by the amount and diversity of training data and availability of annotated images¹¹. In the context of medical imaging, data-sets are often siloed across many institutes, highly unbalanced due to the low incidence of pathology, and often inaccessible due to privacy regulations^{12,13}.

Unlike traditional centralized learning, federated learning (FL)¹⁴ enables multiple parties to collaboratively train a machine learning model without exchanging the underlying data sets. Despite its promising results in medical imaging^{15–23}, the statistical heterogeneity in the data present at the different distributed clients negatively affects the federated performance^{14,18,21,24}. This is especially the case for brain MR scans which are inherently heterogeneous due to e.g., acquisition parameters, different manufactures of medical devices, and different sequences. Recent methods were proposed to tackle data heterogeneity and domain shifts (non-IID), e.g., by not averaging local statistics (SiloBN)²⁵.

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

Figure 1. Our federated unsupervised anomaly segmentation system: we collaboratively train a neural network from multiple institutes (A-D), each with its local dataset. For every training round, each institute sends only its model parameters to a server without sharing private local data. The server then aggregates the parameters of all clients and sends back the updated model. We train the neural network in an unsupervised manner without requiring expansive expert annotation. In doing so, we model the healthy anatomy of the human brain by learning to compress and then reconstruct healthy samples. At inference (Site 1-5), the model sees abnormal samples as input, reconstructs the healthy version of these scans, and then segments the abnormalities given by the residual image. Our main contributions are to disentangle the parameters and leverage global anatomical structure while mitigating domain shifts; and use self-supervision techniques to enforce healthy reconstructions.

Inspired by recent works on disentangled representations²⁶⁻³¹, and the promising findings of anomaly detection in Brain MRI³²⁻³⁷, we propose a federated, unsupervised, domain-agnostic federated method to segment multiple abnormal brain MR findings (multiple sclerosis and tumors). Figure 1 shows an overview of our proposed approach. We learn the healthy normative of a brain by learning how to efficiently compress and encode healthy scans from multiple institutions (A-D), then learn how to reconstruct the data back as close to the original input as possible. At inference time (sites 1-5), the abnormal regions are given by the residual of the abnormal input images and their reconstructed healthy version.

The contributions of this work lie in applying FL concepts to automate the segmentation of abnormalities in brain MRIs. Specifically, we developed FedDis that is able to leverage MR scans from four different sites containing multiple scanner manufacturers (Siemens, Philips), without sharing local data or the need for annotated samples. To mitigate the domain shifts of the different clients, we propose to disentangle the neural network to learn global, shared parameters and local, personalized parameters. This allows the network to leverage mutual anatomical structure information across different institutes and acquisition parameters while keeping appearance/contrast information specific to each client. Furthermore, we show how this property can be enforced with the use of a latent contrastive loss and use self-supervision to enforce

the healthy reconstruction of abnormal input. We extensively evaluated the performance of our method on multiple sites containing real MR images containing multiple scanners and anomalies including multiple sclerosis (MS), glioblastoma (GB), and lower-grade glioma (LGG) and showed superior performance of our method.

Results

Anomaly detection performance

We evaluate the performance of our method on 5 different institutions and datasets (3 public, 2 internal) that include multiple sclerosis lesions (MSLUB, MISBI, MSKRI) and glioblastoma (GBKRI, BraTS). We report the structural similarity index (SSIM) on unseen healthy test data of the participating institutes to evaluate the ability to reconstruct healthy samples, and the area under precision/recall curves (AUPRC) to assess the pathology detection performance. Table 1 reports the aforementioned evaluation metrics of different methods, namely Baur et al.³²: spatial autoencoders trained on KRI; Local: model trained on a single local dataset; FedAvg¹⁴: the federated baseline; SiloBN²⁵: a federated method that tackles data heterogeneity; our proposed method, FedDis; and Data-centralized: model trained with all datasets centralized in a data-lake. To ensure a fair evaluation of the federated disentangled contribution, we use our self-supervision con-

Table 1. Experimental results. (a) shows reconstruction accuracy for healthy unseen tests measured by the structural similarity (SSIM) index. (b) shows area under the precision-recall curves (AUPRC) to assess the anomaly detection performance on different datasets with multiple sclerosis brain tumors.*/** mark statistical significant improvements over the local method and best federated method (FedAvg) respectively (ks-test; $p \leq 0.05$).

Task	Method		SSIM↑ (Healthy) / RI (Average) (%) ↑								
	Client		Lower-bounds		Federated (baselines)		Federated (ours)			Upper-bound	
			Baur et al. ³²	Local	FedAvg ¹⁴	SiloBN ²⁵	FedDis(ours)	-w/o LOL	-w/o SCL	-w/o LCL	Data-centralized
Healthy	Test sets	OASIS(Si)	N/A	0.879	0.870	0.864	0.878	0.877	0.877	0.881	0.903
		ADNI(Si)	N/A	0.867	0.872	0.869	0.875	0.876	0.872	0.880	0.903
		ADNI(Ph)	N/A	0.861	0.876	0.868	0.885	0.885	0.882	0.891	0.911
		KRI(Ph)	N/A	0.807	0.843	0.839	0.840	0.840	0.845	0.850	0.869
		Average	N/A	0.854	0.865	0.860	0.870	0.870	0.869	0.876	0.897
		RI	N/A	0	1.38	0.76	1.87	1.87	1.82	2.58	5.04
Task	Method		AUPRC↑ (Pathology) / RI (Average) (%) ↑								
	Client		Lower-bounds		Federated (baselines)		Federated (ours)			Upper-bound	
			Baur et al. ³²	Local	*FedAvg ¹⁴	SiloBN ²⁵	*FedDis(ours)	*-w/o LOL	*-w/o SCL	**w/o LCL	Data-centralized
Pathology	MSLUB	OASIS(Si)	N/A	0.116	0.102	0.135	0.151	0.154	0.122	0.160	0.136
		ADNI(Si)	N/A	0.070	0.102	0.064	0.164	0.171	0.155	0.175	0.136
		ADNI(Ph)	N/A	0.046	0.102	0.105	0.130	0.135	0.148	0.148	0.136
		KRI(Ph)	0.065	0.047	0.102	0.046	0.146	0.151	0.158	0.169	0.136
		Average	0.065	0.070	0.102	0.088	0.148	0.153	0.146	0.163	0.136
		RI	N/A	0	46.21	25.45	111.83	119.00	108.96	133.69	94.98
	MSISBI	OASIS(Si)	N/A	0.067	0.068	0.077	0.110	0.107	0.055	0.105	0.094
		ADNI(Si)	N/A	0.041	0.068	0.033	0.084	0.084	0.082	0.082	0.094
		ADNI(Ph)	N/A	0.029	0.068	0.055	0.072	0.073	0.070	0.074	0.094
		KRI(Ph)	0.037	0.022	0.068	0.022	0.078	0.077	0.070	0.075	0.094
		Average	0.037	0.040	0.068	0.047	0.086	0.085	0.069	0.084	0.094
		RI	N/A	0	71.07	17.61	116.35	114.47	74.21	111.32	136.48
	MSKRI	OASIS(Si)	N/A	0.304	0.286	0.333	0.273	0.279	0.327	0.263	0.339
		ADNI(Si)	N/A	0.226	0.286	0.222	0.360	0.365	0.190	0.390	0.339
		ADNI(Ph)	N/A	0.086	0.286	0.273	0.283	0.281	0.312	0.304	0.339
		KRI(Ph)	0.130	0.130	0.286	0.199	0.360	0.366	0.348	0.425	0.339
		Average	0.130	0.187	0.286	0.257	0.319	0.323	0.294	0.346	0.339
		RI	N/A	0	53.35	37.67	71.05	73.06	57.77	85.25	81.77
	GBKRI	OASIS(Si)	N/A	0.280	0.285	0.322	0.322	0.322	0.236	0.328	0.274
		ADNI(Si)	N/A	0.241	0.285	0.170	0.353	0.355	0.356	0.362	0.274
		ADNI(Ph)	N/A	0.176	0.285	0.300	0.307	0.306	0.299	0.325	0.274
		KRI(Ph)	0.179	0.172	0.285	0.167	0.305	0.304	0.302	0.293	0.274
		Average	0.179	0.217	0.285	0.240	0.322	0.322	0.298	0.327	0.274
		RI	N/A	0	31.19	10.36	48.10	48.10	37.28	50.52	26.12
	BraTS	OASIS(Si)	N/A	0.333	0.367	0.384	0.411	0.397	0.229	0.387	0.399
		ADNI(Si)	N/A	0.331	0.367	0.250	0.448	0.444	0.424	0.429	0.399
		ADNI(Ph)	N/A	0.313	0.367	0.391	0.377	0.370	0.323	0.373	0.399
		KRI(Ph)	N/A	0.230	0.367	0.259	0.367	0.362	0.313	0.319	0.399
		Average	N/A	0.302	0.367	0.321	0.401	0.393	0.322	0.377	0.399
		RI	N/A	0	21.62	6.38	32.81	30.32	6.79	24.94	32.23

tribution to boost the performance of all baselines, as shown in Figure 3.

As expected, having access to all data samples in a data-lake training achieves the best reconstruction fidelity (SSIM), significantly improving the site results of the local clients with 5.04% on average. Note that, the federated methods improve the reconstruction fidelity of local clients even without sharing local data. Our proposed method, FedDis, achieves the best results among the federated baselines and improves the local/site reconstruction fidelity on average by up to 2.58%. Note that, while clients such as OASIS, with a large amount of training data, benefit marginally from the federated training on the local reconstruction task, clients with less training data such as ADNI-P or KRI benefit from the learned global shape model of FedDis and improve their reconstruction fidelity by 3.37% and 5.06% respectively.

The next sections in the table present the model’s capacity to generalize to unseen sites and segment pathology. Both the

data-centralized and the federated methods improve anomaly detection results over the local clients. SiloBN handles data heterogeneity by averaging only the batch normalization (BN) parameters, while keeping BN statistics, i.e., running mean and variance private. This improves the results for the site with the largest amount of data points (OASIS) outperforming simple federated averaging. However, the rest of participants to the federation are negatively affected, resulting in poor average scores. We hypothesize that SiloBN adapts the BN parameters to the dominating client, hindering the rest of the clients to generalize as well. Our proposed method mitigates the statistical heterogeneity by leveraging the common anatomical information of the human brain and improves the detection of multiple sclerosis on average by up to 110% and 35% over the local and federated averaging, respectively. Similarly, FedDis is able to better detect tumors by up to 41% and 12% on average compared to the local and federated averaging, respectively. The federated paradigm is especially

Figure 2. Qualitative results. The columns show sample segmentation results of each different method while the rows show results on the five different datasets, starting with more difficult cases to cases with good anomaly segmentation results. Our method, FedDis, reduces the amount of false positives and negatives and has a more robust segmentation output for both multiple sclerosis and brain tumors. We highlight qualitative results for OASIS(Si) since it achieves the best local performance.

beneficial for clients performing anomaly detection that also share healthy data from the same institute. Specifically, client KRI improved its local performance from 0.130 up to 0.425 (up to 226%) for MS lesions, and from 0.172 up to 0.305 (up to 77%) for glioblastoma. Surprisingly, our proposed method outperforms the upper-bound, data-centralized model in most cases, indicating that our disentanglement property by leveraging only global information across different sites improves anomaly detection results, even without sharing local data.

Figure 2 gives more insight into the network’s predictions and shows generated segmentation masks of different methods on both MS Lesion and Glioma coming from different testing sites. FedDis reduces the number of false positives and negatives and has a more robust segmentation output for both MS and GB pathologies.

Effect of self-supervision

We train our models with healthy MR scans from different datasets and institutes. However, given the older age of the patients, see Figure 1, multiple MRI scans considered healthy contain hyper-intense regions. The presence of hyper-intense regions in the training set is problematic since the model

will model these hyper-intense regions as part of the healthy anatomy and might be not able to accurately detect pathology, e.g., MS lesions. Based on this observation, we propose to enforce the healthy reconstruction of samples with two self-supervision techniques, as visualized in Fig. ?? . First, to ensure that our training set is indeed healthy, we clean the dataset by painting over values larger than the 98th percentile. Second, we use a strong augmentation technique by drawing rectangles of various sizes and bright intensities over the input sample and force the network to in-paint over these regions. Figure 3 shows AUPRC results for all baselines with and without self-supervision. Our proposed self-supervision technique is applicable to all methods and improves anomaly detection performance. Our method achieves the best anomaly detection scores for both with and without self-supervision setup. Note that, in the absence of self-supervision, our method avoids the reconstruction of anomalies the best, achieving significantly better results than the baselines.

Effect of demographics

Sources of data heterogeneity beyond acquisition parameters and different vendors, e.g., age, gender, or ethnicity,

Figure 3. Fig. 3a shows the precision-recall curve for our proposed method on client OASIS(Si) and Fig. 3b shows different methods for choosing an operating point. We opt for an unsupervised method, where each threshold is selected to output less than 1% false positive rate on their healthy test sets. Fig. 3c shows the influence of self-supervision on performance measured in AUPRC for different methods and datasets. Our contribution increases the anomaly scores consistently on all methods by a significant margin.

Figure 4. Influence of demographics on performance. Fig. 3b shows three different alternatives to choose an operating point for our method. We opt for the unsupervised method where we choose an operating point that produces less than 1% false positive rate on a healthy test. Fig. 4a, 4b, and 4c show the performance of FedDis on MSLUB and BraTS measured in DICE per patient for different gender and age groups, respectively.

might negatively influence the performance of neural networks. To assess the generalizability and clinical applicability of our method, we measured the performance on different gender and age groups for both MS (MSLUB) and brain tumors (BRATS) pathology shown in Figure 4. The patient-wise DICE for male (N=7) and female (N=23) patients was illustrated in Fig. 4a. We found there is no statistically significant difference in the mean DICE performance (0.152 ± 0.142 for men and 0.153 ± 0.139 for women). This observation holds for all baselines. Fig. 4b shows patient-wise DICE for different age groups, patients younger than 40 years (N=20); between 40, and 60 years (N=9) and above 60 (N=1). Interestingly, our method achieved the best DICE score (0.179 ± 0.152) for the age group < 40 years, while having a comparable DICE score of just 0.100 ± 0.091 on patients aged 40 to 60. One possible reason for the performance gap could be the different lesions sizes present in the split: $32.67mm^2$ for younger patients compared to $23.67mm^2$ for the age group older than 40. The one patient over 60 had a very small le-

sion ($6.3mm^2$) and just failed. In contrast, we observed no significant difference for detecting high-grade glioma at all different age groups in Fig. 4c. FedDis achieved a DICE of 0.444 ± 0.194 for the age group less than 40 years (N=9); 0.423 ± 0.207 for patients between 40 and 60 years of age (N=64); and 0.459 ± 0.172 for patients above the age of 60 (N=90).

Effect of disease severity

Indicators or classification of disease severity were not provided through radiology reports for multiple sclerosis. However, we performed an analysis on the effect of the lesion/tumor size on the DICE performance of our method, shown in Figure 5. For brain tumors, we show the performance on the BraTS dataset in detecting low-and-high-grade glioma. Despite the bigger mean lesion size of $452mm^2$ in low-grade glioma (LGG) compared to $345mm^2$ in high-grade glioma (HGG), our network achieves slightly worse performance on detecting LGG with a mean DICE of 0.379 ± 0.169 compared to 0.425 ± 0.193 per patient with HGG. Note that,

Figure 5. Effect of disease severity on performance. On the top, we show the performance (DICE) for different lesions sizes for multiple sclerosis and brain tumors and DICE performance on low- and high-grade glioma. On the bottom row, we show outliers from both multiple sclerosis and brain tumors (small and large lesions).

our algorithm performs better with increasing lesion size. Specifically, 28.85% and 29.60% of the evaluated MR slices had a DICE score above the average 0.23 and 0.39 for MS and brain tumors, respectively. Interestingly, 51.67% and 44.44% of the MR slices contain an insignificant lesion ($< 4mm^2$) or slices where no pathology is present, as can be seen in the top rows of the bottom figures in Figure 5. By removing these slices from the computation, we increase the dice from 0.23 to 0.28 and from 0.39 to 0.41 for MS and brain tumors, respectively. Interestingly, the slices with a mean lesion size above $300mm^2$ achieve a lower dice score for MS. A possible explanation of this behavior is the capability of the networks to reconstruct larger lesions, as visualized in the bottom left plot of Figure 5, and thus leading to poorer anomaly detection.

Effect of disentanglement

Medical data is inherently heterogeneous due to e.g., acquisition parameters, different manufactures of medical devices, local demographics, and rare pathology occurrence. Especially in the case of multi-institutional training, the different sources of heterogeneity can harm the convergence of neural networks and thus limit their clinical applicability. To mitigate the domain shift among the different distributed clients, we disentangle the parameters of the neural network into global, shared- and local, personalized parameters. This allows the network to leverage global information from all clients, e.g., anatomical structure information that should be rather alike across different institutes, while maintaining lo-

cal, client-specific parameters which could encode for e.g., contrast information.

Interestingly, we found out that FedDis with the latent contrastive loss (LCL) works best for detecting tumors while FedDis without LCL achieves better results for detecting MS lesions. This is in contrast to our preliminary works²⁶, where we show that LCL is crucial for improving the anomaly detection scores in both MS lesions and tumors. It is worth mentioning that the only differences between the two works; i) using shallower but wider network architecture to improve the reconstruction fidelity, ii) using self-supervision to enforce healthy reconstructions and iii) adding an additional participating client, namely KRI. This finding suggests that having powerful auto-encoders, that are able to better explore and reconstruct the data, yields similar anomaly detection scores regardless of the additional constraints.

To showcase the effectiveness of the disentanglement, we illustrate in 6a a 2D visualization of the global (shape) and local (appearance) latent embeddings of healthy, unseen samples of the clients that participated in the federated training. Note that, the shape and appearance embeddings are far away from each other, with shape encodings following a similar distribution, while the appearance representations are well separated. To further analyze the latent space, we show four samples from the four datasets that are in close proximity in the shape manifold (0.5 radius). The samples show structural similarity e.g., the shape of the skull, or ventricle size but vary in appearance, e.g. intensity/contrast.

Figure 6. 6a. Global (shape) vs. local (appearance): a visualization of latent embedding using UMAP³⁸. We show on the left the global and local embeddings of healthy, unseen test samples of all four institutes that participated in the federating training. The plot shows different clusters of appearance representations, indicating domain shifts among the institutes, e.g., given by intra- and inter-scanner variations. On the other hand, the shape representations of all datasets cluster together, highlighting the common anatomical structure among different scanners and acquisition parameters. On the right side of the figure, we show that image samples from the different institutes that are in close proximity in the shape-space share similar anatomical features, e.g., the shape of the skull, or ventricle size. Interestingly, the appearance representations of the same samples are located in different appearance clusters, indicating local variations, e.g., in contrast. 6b shows the effect of the appearance parameters. In the two different examples (MS GB) we show from left to right: the original input image; the reconstructed image without using the encoded appearance representation; and the reconstructed image using both shape and appearance embeddings. The added appearance parameters have a styling effect on the reconstruction, making it more similar in terms of contrast to the original input image.

This observation correlates with the appearance embeddings of the same samples belonging to different appearance clusters. 6b show sample reconstructions of our proposed method with and without using the appearance representation. While both reconstructions seem to capture the global shape of the input, adding the local parameters styles the reconstruction to resemble the original input in terms of contrast, indicating that the local parameters encode indeed appearance information.

Discussion

Our results showed that the proposed automated, unsupervised neural network-based method can identify anomalies and segment critical findings including MS lesions and tumors on

brain MRI scans from multi-vendor imaging systems coming from different sites. Although DL methods have been used to detect multiple abnormal findings on brain MRI, available solutions for detection and assessment of disease progression have been limited to a single dataset and/or require expensive labeling for supervised training. In this work, we proposed a privacy-aware, unsupervised method to detect abnormal MR scans from multiple institutes and datasets to train a single global model without sharing local sensitive data. We do so, by deploying a neural network directly to the clients, train locally, and aggregate only the updated models of the local institutes. Furthermore, we proposed to mitigate the domain shifts among distributed clients by disentangling shared global features, e.g., anatomical information, from

personalized local features, e.g., appearance. We evaluated our proposed method on multiple datasets containing real pathology data and showed a superior performance over a locally trained model by 99.74% for MS and 40.45% for tumors.

Clinical relevance. The commonly used approach in clinical practice to find and assess brain MR findings is a naive thresholding-based classifier. This approach requires no training data and selects hyper-intense regions in the input images based on a given threshold. We optimize the operating point using all labeled samples for each of the datasets, achieving thus the upper-bound performance for this approach, with a mean DICE/patient of 0.173 ± 0.177 for MS lesions and 0.327 ± 0.243 for brain tumors. Even though we do not use any labeled test data for selecting an operating point, our proposed method shows superior performance achieving 0.191 ± 0.131 (relative improvement of 10.09%) and 0.360 ± 0.189 (relative improvement of 10.40%) DICE/patient for MS lesions and brain tumors, respectively.

Scalability. In our experiments, we leveraged FLAIR images from four institutes and various 3 Tesla scanner manufacturers. However, in clinical practice, institutes may use other MR sequences for detecting pathology, e.g., T1-weighted images, or diffusion-weighted images (DWI). To assess the scalability of our approach we added four more clients containing T1w MR scans from different 1.5 and 3T scanners and evaluated the diagnostic performance of the different networks. Our method, FedDis, was not affected by the domain shift and achieved a similar performance when the four T1w clients were added. However, FedAvg and SiloBN were severely affected by the increased data heterogeneity due to different MR sequences, showing a drastic drop in their diagnostic performance by 33%, and 36%, respectively. Usually, clinical institutes acquire different MR sequences of the same patient in one scan, e.g., T1w and T2w images. While these sequences vary in tissue representation and contrast, they both capture the same underlying anatomical structure. Based on this observation, we could formulate the shape consistency loss to enforce similar shape embeddings for different co-registered sequences. To test this hypothesis, we adapt our internal client, KRI, to enforce the shape consistency based on T1w/Flair sequence pairs. The other three clients in the federation compute the shape consistency loss using intensity-augmented versions of the original FLAIR images, as described in the methods section. Interestingly, early results show a slight improvement in the detection performance when using pairs of sequences for the shape consistency loss. However, more work is required to analyze the slight shape difference between T1w and Flair sequences and their impact on the performance.

Thresholding. To analyze the diagnostic performance of the different models, we reported the area under the precision-recall curves (AURPC) in Table 1. The question of how to balance between precision and recall and choose an operating

point for a model remains open. Similar to ³⁹, we followed the approach proposed by ⁴⁰ where we opt for an institute-specific threshold that results in a very few mistakes ($< 1\%$ false positive rate) on their corresponding healthy test data. We compared the unsupervised choice of the operating point with two different methods that require annotated samples of the test client: choosing a threshold on a random subset containing 15% of annotated samples and optimizing the best operating point for the whole dataset, i.e., upper-bound DICE. Interestingly, the unsupervised approach achieved comparable performance to the one using 15% of the annotated samples and showed a slight drop in the performance compared to the upper-bound DICE. This suggests the effectiveness of the unsupervised approach⁴⁰.

Complexity. For implementing our proposed method, FedDis, we opted to hold the same amount of shared parameters to be able to capture the global information and add a few parameters for capturing the appearance/local information. However, if we choose to keep the same network complexity and reduce the amount of parameters that are shared globally, our network still achieves a relative improvement of 48%/31.73% and 9.63%/8.91% for MS/brain tumors over the local and federated averaging, respectively. In our experiments, we opt for the standard auto-encoder architecture³². For future work, we plan to investigate more complex architectures for the reconstruction of healthy samples, such as Gaussian mixture variational auto-encoders^{41,42}.

Privacy Concerns. Even though the federated paradigm reduces the privacy risks by not explicitly sharing local data, recent works^{43,44} demonstrated that sharing model updates makes FL vulnerable to inference attacks, i.e., data representation leakage from gradients being the essential cause of privacy leakage. To mitigate this issue, recent works⁴⁵ have been proposed to, e.g., encrypt gradient updates from clients to the server or withhold individual information from global statistics using differential privacy. These works are complementary to our approach and can be integrated into our pipeline to mitigate privacy risks.

Methods

The main concept behind our federated unsupervised anomaly segmentation framework, depicted in Figure 1, is to model the distribution of healthy anatomy by learning how to efficiently compress and encode healthy brain scans from multiple institutions, then learn how to reconstruct the data back as close to the original input as possible. This enables the detection of pathology from faulty reconstructions of anomalous samples. We first formally introduce the problem and define the federated unsupervised anomaly segmentation setup. Then, we present FedDis and elaborate a loss to enforce the disentanglement.

Problem formulation Given M clients $\mathcal{C}_j$ with local dataset $\mathcal{D}_j \in \mathbb{R}^{H \times W \times N_j}$ consisting of N_j healthy brain MR scans

$x \in \mathbb{R}^{H \times W}$, our objective is to train a global model $f_{\theta^G}(\cdot)$, leveraging the healthy brain scans from multiple institutes without sharing local data, to detect and segment the pathology $r_q \in \mathbb{R}^{H \times W}$ for a given query brain scan x_q ;

$$r_q = \max(0, x_q - x_{Rec}) > \tau, \text{ with } x_{Rec} = f_{\theta^G}(x_q), \quad (1)$$

where τ is a certain threshold.

FedDis: Federated Disentangled Representation Learning for Unsupervised Brain Anomaly Detection. FL aims to collaboratively learn a global model $f_{\theta^G}(\cdot)$, without centralizing or sharing training data. At each communication round, the local clients are initialized with the global weights and trained locally on their own data sets for a fixed number of epochs to minimize following objective function:

$$\min_{\theta^G} \mathcal{L}(\mathcal{D}; \theta^G) := \sum_{j=1}^M w_j \mathcal{L}_{Rec}(\mathcal{D}_j; \theta^{C_j}), \quad (2)$$

where the learned local parameters θ^{C_j} are aggregated to a new global model: $\theta^G \leftarrow \sum_{j=1}^M w_j \theta^{C_j}$, where $w_j = \frac{N_j}{\sum_{j=1}^M N_j}$ is the respective weight coefficient. A popular architecture to learn efficient data representation in an unsupervised manner are convolutional auto-encoders (AE)⁴⁶ where an encoder is trained to compress x to a latent representation $z \in \mathbb{R}^d$, from which a decoder attempts to reconstruct the original by minimizing following objective:

$$\arg \min_{\theta^{C_j}} \sum_x \mathcal{L}_{Rec}(x, x_{Rec}), \text{ with } x_{Rec} = f_{\theta^{C_j}}(x), \quad (3)$$

where a common choice for the reconstruction loss is the mean absolute error: $\mathcal{L}_{Rec}(x, x_{Rec}) = \frac{1}{N} \sum_{i=0}^N |x - x_{Rec}|$.

Disentanglement To mitigate the statistical heterogeneity, but leverage the shared structural anatomical information among the distributed clients, we propose to disentangle the model parameters θ into shape θ_S and appearance θ_A and only share the former in the federation. Thus, after every communication round, the global model parameter is updated as follows: $\theta^G \leftarrow \sum_{j=1}^M w_j \theta_S^{C_j}$, where $\theta_S^{C_j}$ is the shape parameter at the client C_j .

To further enforce the disentanglement, we introduce following losses: (i) shape consistency loss (SCL): shape embeddings of a given brain scan should be similar under different intensity augmentations e.g., changes in brightness/contrast, random gamma shifts, or corresponding pairs of different MR sequences - we choose the latter two; and (ii) latent orthogonality loss (LOL): latent representations for shape and appearance should be orthogonal to each other. Thus, we define the latent contrastive loss (LCL) as

$$\mathcal{L}_{LCL}(z_A, z_S, z_{\gamma S}; \theta_S; \theta_A) = \underbrace{\beta_S \mathcal{L}_{CS}(z_S, z_{\gamma S}; \theta_S)}_{\mathcal{L}_{SCL}} + \underbrace{\beta_L (1 - \mathcal{L}_{CS}(z_A, z_S; \theta_A))}_{\mathcal{L}_{LOL}}, \quad (4)$$

where z_A, z_S and $z_{\gamma S}$ are the latent representation for the appearance, shape and shape of the gamma shifted image, respectively; β_S and β_L are used to weigh the two contrasting loss terms, and $\mathcal{L}_{CS}(x_1, x_2) = \cos(x_1, x_2)$ is the cosine similarity.

The overall objective is given by: $\mathcal{L}_{Rec}(x, x_{Rec}) + \alpha \mathcal{L}_{LCL}(z_A, z_S, z_{\gamma S})$, with regularization weight α . Finally, the anomaly segmentation is given by the binarization of the residual $r = \max(0, x_q - x_{Rec}) > \tau$.

Implementation details and hyper-parameters. For FedDis, we followed³² by using an AE consisting of 3 layers of 5×5 convolutions with filters ranging from 64 to 128 with the spatial bottleneck $z \in \mathbb{R}^{16 \times 16 \times 128}$ for encoding the global shape parameters and 3 layers of 5×5 convolutions with filters ranging from 16 to 32 with the spatial bottleneck $z \in \mathbb{R}^{16 \times 16 \times 32}$ to capture the local appearance. Each convolution layer is followed by batch normalization and Leaky ReLU activation layers. To avoid over-fitting we add a dropout of 0.2 to the last convolutional layer for all methods. We trained the models for 50 rounds, each with 5 local epochs with a batch size of 8. We used an ADAM⁴⁷ with a learning rate of 1^{-4} and exponential decay of 0.97. We set the loss weight α to 0.2 after a grid search with $\alpha \in 0.2, 0.5, 0.75, 1$. β_S and β_L pairs correspond to different variants of our method: FedDis: (1,1); FeDis-w/o LOL: (1,0); FedDis-w/o SCL: (0,1) and FedDis-w/o LCL: (0,0). We used the MONAI⁴⁸ implementation for the gamma shift, with random values in the interval $[0.5, 2]$. The objective loss imposes two contrasting tasks: i) learn to eliminate masking and hyper-intensities from the reconstruction and ii) have a similar shape embedding for two intensity-augmented sequences (SCL). Thus, we consider it is useful to inject the regularization loss (LCL) at a later time-point in the optimization, after the network learns to reconstruct healthy samples. We did a hyper-parameter search for the time-point $t \in 0, 5, 25, 45$ in which to inject the contrastive loss and found that injecting the latent contrastive loss at round 25 in the federated training yields the best results. For the self-supervision augmentation, we first clean the images by painting over high-intensity values (> 98 th percentile) with the mean intensity value of the brain slice. Second, we augment the original images with up to 3 rectangles of random size $s \in [6, 500]$ with the width $w \in [3, 20]$ and height $h \in [w - w/3, w + w/3]$. The position of the rectangles is random with the coordinates of the top left corner given by: $x \in [20, 90]$ and $y \in [30, 90]$. The intensity of the painted rectangles is uniform, with the random brightness value in the interval given by the value of 99th percentile of the brain slice and 1. We used this augmented image as an input of our networks for training and the cleaned image as the ground truth.

Datasets. An overview of the datasets used for training and evaluation is shown in Table 1. We used two publicly available brain MR datasets (OASIS-3 and ADNI-3) and one internal database KRI for training.

OASIS-3¹ dataset containing 1089 patients collected over 30 years. Participants (aged 42-95) include 609 cognitively normal adults and 489 individuals at various stages of cognitive decline. We used all 633 FLAIR sequences of 450 healthy patients for training.

ADNI² The Alzheimer’s Disease Neuroimaging Initiative (ADNI) is a longitudinal multicenter study designed for the early detection of Alzheimer’s disease. We used the healthy control group and selected all 507 FLAIR MR scans of 118 patients acquired with multiple Siemens scanners for our training client, **ADNI-S**. Additionally, we selected all 228 FLAIR MR scans of 55 healthy patients acquired with multiple Philips scanners for our training client, **ADNI-P**.

KRI is our internal dataset containing 163 Flair, T2w, and T1w co-registered scans acquired with a 3T Philips Achieva scanner.

To assess the diagnostic performance of our network, we used two publicly available MS lesion datasets (MSLUB, MSISBI); an in-house database with MS and Glioblastoma (MSKRI, GBKRI), and a publicly available dataset containing brain tumors (BraTS).

MSLUB³ contains Flair, T2w, and T1w scans of 30 patients acquired with a 3T Siemens Magnetom Trio MR scanner at the University Medical Center Ljubljana (UMCL). Segmentation masks for MS are provided as a consensus of three experts.

MSISBI⁴ contains 21 Flair, T2w, T1w, and PD scans of 5 patients acquired with a 3T Philips Medical Systems scanner in Best, Netherlands and is the publicly available training set of the 2015 Longitudinal MS lesion segmentation challenge.

MSKRI and **GBKRI** are our internal dataset containing 48/94 Flair, T2w, and T1w scans acquired with a 3T Philips Achieva scanner. MS lesion and glioblastoma (whole tumor) segmentation masks are provided by expert neuro-radiologists.

BRATS 2018 dataset contains Flair, T2w, T1w, and T1Gd (post-contrast) scans of 285 patients with low-(N=122) and high-grade tumors(N=163). Segmentation masks are provided manually by one to four raters and approved by experienced neuro-radiologist. For our evaluations, we consider the whole tumor containing the GD-enhancing tumor, the peritumoral edema, and the necrotic and non-enhancing tumor core.

Pre- and post-processing. All scans have been registered to the SRI24 atlas template space⁸ to ensure all data share the same volume size and orientation. Subsequently, the scans have been skull-stripped with ROBEX⁴⁹ and normalized to the [0,1] range. We used the axial mid-line with a size of 128×128px for training and evaluated our methods patient-wise on whole volumes containing slices with visible tissue information. For post-processing, we use prior knowledge and keep only positive residuals, as these lesions are known to be fully hyper-intense in FLAIR images. Further, we apply median filtering of size 3 to remove small outliers and obtain a more continuous signal. We use the resulted heat maps to generate AUPRC for the test sets containing pathology.

Finally, we choose an operating point to binarize the results and compute the DICE score per patient. We choose the operating point τ in an unsupervised manner⁴⁰ by choosing the lowest threshold that delivers a false positive rate lower than 1% on the healthy test set of each client.

Evaluation metrics. To measure the anomaly segmentation performance and compare different models, we report the area under the precision $TP/(TP + FP)$ - recall $TP/(TP + FN)$ curves (AUPRC), with TP , FP , and FN being true positives, false positives and false negatives, respectively. We also report DICE scores per patient given by $2TP/(2TP + FP + FN)$. We report the relative improvement (RI) of a over b as $(a - b)/b$ and use the structure similarity SSIM⁵⁰ to measure the reconstruction fidelity. Finally, we used the Kolmogorov–Smirnov test: $KS(F(x), G(x))$ to measure statistical significant differences between models F and G with x being patient-wise DICE values. The null hypothesis is that $F(x) \leq G(x)$ for all x ; and the alternative is that $F(x) > G(x)$ for at least one x , with a p-value lower than 0.05 suggesting a stronger evidence in favor of the alternative hypothesis.

Data Availability

Most of the data-sets used in this study are publicly available and can be downloaded. The OASIS dataset is available at <https://www.oasis-brains.org>; the ADNI-S and ADNI-P datasets are available at <http://adni.loni.usc.edu/data-samples/access-data/>; the MSLUB dataset is available at <http://lit.fe.uni-lj.si/tools.php?lang=eng>; the MSISBI dataset is available at <https://smart-stats-tools.org/lesion-challenge-2015>; and the BRATS 2018 dataset is available at <https://www.med.upenn.edu/sbia/brats2018/data.html>.

Code Availability

The code will be made publicly available upon acceptance at <https://github.com/albarqounilab/FedDis>.

References

1. LaMontagne, P. J. *et al.* Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medRxiv* (2019).
2. Weiner, M., Veitch, D. & PS, A. The alzheimer’s disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. *Alzheimers Dement.* (2016).
3. Lesjak, Z. *et al.* A novel public mr image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus. *Neuroinformatics* **16** (2018).
4. Carass, A. *et al.* Longitudinal multiple sclerosis lesion segmentation data resource. *Data Brief* **12**, 346–350 (2017).
5. Menze, B. *et al.* The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans Med Imaging* (2015).

- 578 6. Bakas, S. *et al.* Advancing the cancer genome atlas
579 glioma mri collections with expert segmentation labels
580 and radiomic features. *Sci Data* (2017).
- 581 7. Bakas, S. *et al.* Identifying the best machine learning
582 algorithms for brain tumor segmentation, progression
583 assessment, and overall survival prediction in the brats
584 challenge (2019). [1811.02629](#).
- 585 8. Rohlfing, T., Zahr, N. M., Sullivan, E. V. & Pfefferbaum,
586 A. The sri24 multichannel atlas of normal adult human
587 brain structure. *Hum. brain mapping* **31**, 798–819 (2010).
- 588 9. Klawiter, E. C. Current and new directions in mri in mul-
589 tiple sclerosis. *Continuum (Minneapolis, Minn.)* (2013).
- 590 10. Soltaninejad, M. *et al.* Automated brain tumour detection
591 and segmentation using superpixel-based extremely ran-
592 domized trees in flair mri. *Int. J. Comput. Assist. Radiol.*
593 *Surg.* (2017).
- 594 11. Sun, C., Shrivastava, A., Singh, S. & Gupta, A. Revisiting
595 unreasonable effectiveness of data in deep learning era.
596 In *Proceedings of the IEEE international conference on*
597 *computer vision*, 843–852 (2017).
- 598 12. Rieke, N. *et al.* The future of digital health with federated
599 learning. *NPJ digital medicine* **3**, 1–7 (2020).
- 600 13. Kaissis, G. A., Makowski, M. R., Rückert, D. & Braren,
601 R. F. Secure, privacy-preserving and federated machine
602 learning in medical imaging. *Nat. Mach. Intell.* **2**, 305–
603 311 (2020).
- 604 14. McMahan, B., Moore, E., Ramage, D., Hampson, S. &
605 y Arcas, B. A. Communication-efficient learning of deep
606 networks from decentralized data. In *Artificial Intelli-*
607 *gence and Statistics*, 1273–1282 (PMLR, 2017).
- 608 15. Collaborative learning without sharing data. *Nat. Mach.*
609 *Intell.* **3**, 459–459, DOI: [10.1038/s42256-021-00364-5](#)
610 (2021).
- 611 16. Dou, Q. *et al.* Federated deep learning for detecting
612 covid-19 lung abnormalities in ct: a privacy-preserving
613 multinational validation study. *NPJ digital medicine* **4**,
614 1–11 (2021).
- 615 17. Sheller, M. J. *et al.* Federated learning in medicine: facil-
616 itating multi-institutional collaborations without sharing
617 patient data. *Sci. reports* **10**, 1–12 (2020).
- 618 18. Li, X. *et al.* Multi-site fmri analysis using privacy-
619 preserving federated learning and domain adaptation:
620 Abide results. *Med. Image Analysis* **65**, 101765 (2020).
- 621 19. Li, D., Kar, A., Ravikumar, N., Frangi, A. F. & Fidler,
622 S. Federated simulation for medical imaging. In *Inter-*
623 *national Conference on Medical Image Computing*
624 *and Computer-Assisted Intervention*, 159–168 (Springer,
625 2020).
- 626 20. Sarma, K. V. *et al.* Federated learning improves site
627 performance in multicenter deep learning without data
628 sharing. *J. Am. Med. Informatics Assoc.* (2021).
21. Yang, D. *et al.* Federated semi-supervised learning for
covid region segmentation in chest ct using multi-national
data from china, italy, japan. *Med. Image Analysis* 101992
(2021).
22. Albarqouni, S. *et al.* *Domain Adaptation and Repre-*
sensation Transfer, and Distributed and Collaborative
Learning: Second MICCAI Workshop, DART 2020, and
First MICCAI Workshop, DCL 2020, Held in Conjunc-
tion with MICCAI 2020, Lima, Peru, October 4-8, 2020,
Proceedings, vol. 12444 (Springer Nature, 2020).
23. Bdair, T., Navab, N. & Albarqouni, S. Fedperl: Semi-
supervised peer learning for skin lesion classification. In
International Conference on Medical Image Computing
and Computer-Assisted Intervention (Springer, 2021).
24. Konečný, J., McMahan, H. B., Ramage, D. & Richtarik,
P. Federated optimization: Distributed machine learning
for on-device intelligence (2016).
25. Andreux, M., du Terrail, J. O., Beguier, C. & Tramel,
E. W. Siloed federated learning for multi-centric
histopathology datasets. In *Domain Adaptation and Rep-*
resentation Transfer, and Distributed and Collaborative
Learning, 129–139 (Springer, 2020).
26. Bercea, C. I., Wiestler, B., Rueckert, D. & Albarqouni,
S. Feddis: Disentangled federated learning for unsu-
pervised brain pathology segmentation. *arXiv preprint*
arXiv:2103.03705 (2021).
27. Chartsias, A. *et al.* Disentangled representation learn-
ing in cardiac image analysis. *Med. Image Analysis* **58**,
101535 (2019).
28. Locatello, F. *et al.* Challenging common assumptions in
the unsupervised learning of disentangled representations.
In *International Conference on Machine Learning* (2019).
29. Sarhan, M. H., Navab, N., Eslami, A. & Albarqouni, S.
Fairness by learning orthogonal disentangled represen-
tations. In *European Conference on Computer Vision*,
746–761 (Springer, 2020).
30. Shu, Z. *et al.* Deforming autoencoders: Unsupervised
disentangling of shape and appearance. In *Proceedings*
of the European conference on computer vision (ECCV),
650–665 (2018).
31. Qin, C. *et al.* Unsupervised deformable registration for
multi-modal images via disentangled representations. In
International Conference on Information Processing in
Medical Imaging, 249–261 (Springer, 2019).
32. Baur, C., Denner, S., Wiestler, B., Navab, N. & Albar-
qouni, S. Autoencoders for unsupervised anomaly seg-
mentation in brain mr images: A comparative study. *Med.*
Image Analysis 101952 (2021).
33. Chen, X., You, S., Tezcan, K. C. & Konukoglu, E. Un-
supervised lesion detection via image restoration with a
normative prior. *Med. image analysis* **64**, 101713 (2020).

680	34. Pinaya, W. H. L. <i>et al.</i> Unsupervised brain anomaly detection and segmentation with transformers. <i>arXiv preprint arXiv:2102.11650</i> (2021).	with publicly available methods. <i>IEEE transactions on medical imaging</i> 30 , 1617–1634 (2011).	731
681			732
682			
683	35. van Hespen, K. M. <i>et al.</i> An anomaly detection approach to identify chronic brain infarcts on mri. <i>Sci. Reports</i> 11 , 1–10 (2021).	50. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. <i>IEEE transactions on image processing</i> 13 , 600–612 (2004).	733
684			734
685			735
686	36. Baur, C. <i>et al.</i> Modeling healthy anatomy with artificial intelligence for unsupervised anomaly detection in brain mri. <i>Radiol. Artif. Intell.</i> 3 , e190169 (2021).		736
687			
688			
689	37. Nael, K. <i>et al.</i> Automated detection of critical findings in multi-parametric brain mri using a system of 3d neural networks. <i>Sci. reports</i> 11 , 1–10 (2021).	Acknowledgments	737
690		The authors would like to thank their clinical partners at Klinikum rechts der Isar, Munich, for generously providing their data.	738
691			739
692	38. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. <i>arXiv preprint arXiv:1802.03426</i> (2018).		740
693			
694			
695	39. Heer, M., Postels, J., Chen, X., Konukoglu, E. & Albarqouni, S. The ood blind spot of unsupervised anomaly detection. In <i>Medical Imaging with Deep Learning</i> (2021).	Author contributions statement	741
696		C.B. contributed to Methodology, Software, Formal analysis, Investigation, Visualization, and Writing - original draft. B.W. contributed to Data curation, Resources, and and Writing - review editing. D.R contributed to Supervision, and Writing - review editing. S.A. contributed to Supervision, Conceptualization, Methodology, Investigation, Writing - review editing, Resources, and Project administration. All authors proof-read and accepted the final version of the paper.	742
697			743
698	40. Konukoglu, E., Glocker, B., Initiative, A. D. N. <i>et al.</i> Reconstructing subject-specific effect maps. <i>NeuroImage</i> 181 , 521–538 (2018).		744
699			745
700			746
701	41. Dilokthanakul, N. <i>et al.</i> Deep unsupervised clustering with gaussian mixture variational autoencoders. <i>arXiv preprint arXiv:1611.02648</i> (2016).		747
702			748
703			749
704	42. You, S., Tezcan, K. C., Chen, X. & Konukoglu, E. Unsupervised lesion detection via image restoration with a normative prior. In Cardoso, M. J. <i>et al.</i> (eds.) <i>Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning</i> , vol. 102 of <i>Proceedings of Machine Learning Research</i> , 540–556 (PMLR, 2019).	Additional information	750
705		Competing Interests Statement. The authors declare that they have no known competing interests that could have appeared to influence the work reported in this paper.	751
706		Correspondence should be addressed to S.A.	752
707			753
708			754
709			
710	43. Xie, C., Huang, K., Chen, P.-Y. & Li, B. Dba: Distributed backdoor attacks against federated learning. In <i>International Conference on Learning Representations</i> (2019).		
711			
712			
713	44. Lyu, L. <i>et al.</i> Privacy and robustness in federated learning: Attacks and defenses. <i>arXiv preprint arXiv:2012.06337</i> (2020).		
714			
715			
716	45. Sun, J. <i>et al.</i> Soteria: Provable defense against privacy leakage in federated learning from representation perspective. In <i>Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition</i> , 9311–9319 (2021).		
717			
718			
719			
720			
721	46. Vincent, P. <i>et al.</i> Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. <i>J. machine learning research</i> 11 (2010).		
722			
723			
724			
725	47. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. <i>arXiv preprint arXiv:1412.6980</i> (2014).		
726			
727	48. Consortium, T. M. Project monai, DOI: 10.5281/zenodo.4323059 (2020).		
728			
729	49. Iglesias, J. E., Liu, C.-Y., Thompson, P. M. & Tu, Z. Robust brain extraction across datasets and comparison		
730			