Computerized Medical Imaging and Graphics 85 (2020) 101770

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Investigating the impact of supervoxel segmentation for unsupervised abnormal brain asymmetry detection

Samuel B. Martins [a,b,c,*], Alexandru C. Telea [d], Alexandre X. Falcão [a]

[a] Laboratory of Image Data Science (LIDS), Institute of Computing, University of Campinas, Brazil
[b] Bernoulli Institute, University of Groningen, The Netherlands
[c] Federal Institute of São Paulo, Campinas, Brazil
[d] Department of Information and Computing Sciences, Utrecht University, The Netherlands

ARTICLE INFO

Keywords:
Abnormal brain asymmetry
Anomaly detection
Supervoxel segmentation
One-class classification
MR images of the brain

ABSTRACT

Several brain disorders are associated with abnormal brain asymmetries (asymmetric anomalies). Several computer-based methods aim to detect such anomalies automatically. Recent advances in this area use automatic unsupervised techniques that extract pairs of symmetric supervoxels in the hemispheres, model normal brain asymmetries for each pair from healthy subjects, and treat outliers as anomalies. Yet, there is no deep *understanding* of the impact of the supervoxel segmentation quality for abnormal asymmetry detection, especially for small anomalies, nor of the added value of using a specialized model for each supervoxel pair instead of a single global appearance model. We aim to answer these questions by a detailed evaluation of different scenarios for supervoxel segmentation and classification for detecting abnormal brain asymmetries. Experimental results on 3D MR-T1 brain images of stroke patients confirm the importance of high-quality supervoxels fit anomalies and the use of a specific classifier for each supervoxel. Next, we present a refinement of the detection method that reduces the number of false-positive supervoxels, thereby making the detection method easier to use for visual inspection and analysis of the found anomalies.

1. Introduction

Magnetic resonance imaging (MRI) is usually the standard image modality for structural brain analysis, as it provides detailed 3D images with high spatial resolution and high contrast for soft tissues (Akkus et al., 2017). Quantitative analysis of MR brain images has been used extensively for the characterization of brain disorders, such as stroke, tumors, and multiple sclerosis. Such methods rely on delineating objects of interest — (sub)cortical structures or lesions — trying to solve detection and segmentation simultaneously for tasks such as quantitative lesion assessment (*e.g.*, volume), surgical planning, treatment assessment, and overall anatomic understanding (Kamnitsas et al., 2017; Chen et al., 2018b; Soltaninejad et al., 2017).

The simplest strategy to detect brain anomalies consists of a visual slice-by-slice inspection by one or multiple specialists. This process is very time-consuming and even impracticable when large data amounts need to be processed, and it is also inaccurate due to human errors. Continuous efforts have been made for automatic anomaly detection that delineates anomalies with accuracy close to that of human experts. However, this goal is challenging and complex due to the large variability in shape, size, and location among different anomalies, even when the same disease causes these (see, *e.g.*, Fig. 1).

Recently, a method called Supervoxel-based Abnormal Asymmetry Detection (SAAD) (Martins et al., 2019a) was proposed to detect abnormal asymmetries in MR brain images. Unlike many methods that focus on detecting lesions specific to some disease, or morphology, SAAD aims to detect generic lesions focusing on their asymmetry. This is especially interesting since many neurological diseases are associated with abnormal brain asymmetries (Wang et al., 2001). By using a supervoxel segmentation, combined with a one-class per-supervoxel classifier, SAAD claims to obtain higher detection accuracy even for small lesions compared to state-of-the-art detection methods as deep generative neural networks. However, two key questions are still open for SAAD: (i) what is the impact of supervoxel segmentation in SAAD on the quality of the abnormal asymmetry detection? (ii) why use a specialized one-class classifier for each supervoxel instead of a global classifier?

In this paper, we answer these questions by a detailed evaluation of different scenarios for supervoxel segmentation and classification to detect abnormal asymmetries in MR-T1 brain images using the SAAD

* Corresponding author at: Laboratory of Image Data Science (LIDS), Institute of Computing, University of Campinas, Brazil.

https://doi.org/10.1016/j.compmedimag.2020.101770
Received 6 November 2019; Received in revised form 27 July 2020; Accepted 31 July 2020
Available online 19 August 2020

method. Experimental results on 3D MR-T1 brain images with asymmetric stroke lesions confirm the importance of a high-quality fit of supervoxels to lesions and the use of a specific classifier for each supervoxel. Using these insights, we also show how we can improve the detection accuracy quantitatively compared to the original results of SAAD.

The remainder of this paper is organized as follows. Section 2 discusses related work on automatic brain lesion detection and segmentation. Section 3 introduces SAAD with a focus on its supervoxel segmentation step. Section 4 describes the experimental setup we used to analyze how SAAD's performance depends on the supervoxel segmentation and classification scheme used. Section 5 presents and discusses the obtained results. Section 6 concludes the paper.

2. Related work

Broadly speaking, automatic brain lesion detection/segmentation methods can be grouped into five classes. From the least to the most versatile, these are as follows:

Atlas-based methods. These methods use the *a priori* knowledge about the object's shapes in a training atlas set registered on a standard template, where each atlas consists of a source 3D image and its corresponding 3D label image with the mask of each 3D object of interest (Martins et al., 2019b; Aljabar et al., 2009; Lötjönen et al., 2010; Manjón and Coupé, 2016; González-Villà et al., 2019). Shape-constraints are often encoded either on a probabilistic atlas (Martins et al., 2019b) (each voxel has a prior probability of belonging to a given object) or by combining all segmentation masks by label fusion techniques (Aljabar et al., 2009). Although atlas methods show impressive segmentation results for healthy tissues (Martins et al., 2019b; Manjón and Coupé, 2016), they fail to delineate anomalies, especially given the latter's arbitrary shapes and locations (see, *e.g.*, Fig. 1).

Supervised learning with hand-crafted features. These methods use different classifiers trained from various *hand-crafted* image features (*e.g.*, edge detectors and texture features) to delineate anomalies by classifying voxels or regions of the target image (Soltaninejad et al., 2017; Goetz et al., 2014; Geremia et al., 2011; Pinto et al., 2015). Such methods usually do not generalize well since the used features have limited representation capability considering the significant variation of the lesions' appearances. Moreover, these methods work well only for detecting anomalies related to diseases present in the training set.

Discriminative deep learning. These techniques have emerged as a powerful alternative to the previous class of methods, given their ability to learn highly discriminative features for a particular task. In particular, convolutional neural networks (LeCun et al., 2015) have become a mainstay of the computer vision community due to breakthrough performance in several applications (Pouyanfar et al., 2018) as compared to approaches using hand-crafted features. Deep learning has gained popularity in medical image analysis as well (Vasilakos et al., 2016; Kooi et al., 2017; Havaei et al., 2017; Aslani et al., 2018). Such methods learn deep feature representations (*e.g.*, convolutional features) in a data-driven way without any kind of feature engineering being required. Yet, there are some limitations:

(i) They require a large number of training images that must be previously annotated by specialists (*e.g.*, lesion segmentation masks);
(ii) Typically require weight fine-tuning (retraining) when used for a new set of images due to image variability across scanners and acquisition protocols;
(iii) They are only designed for the anomalies found in the training set, just as the supervised methods outlined before; and
(iv) The success of such methods on new images is limited by the absence of large, high-quality, annotated training sets for most medical image analysis problems (Akkus et al., 2017).

Deep generative neural networks. Also known as Encoder-Decoder Neural Networks (EDNNs), or autoencoders, these methods have been used for unsupervised anomaly detection by modeling the distribution of healthy brain tissues and next detecting outliers as anomalies. The underlying hypothesis is that this model can reconstruct healthy brain anatomies while failing to reconstruct anomalies in images with some disorder. EDDNs learn to reconstruct training images from healthy individuals *only* by first compressing (encoding) them into a low-dimensional representation (latent features) and then decompressing that representation to minimize the reconstruction error between the input data and its reconstruction. Some methods (Sato et al., 2018; Baur et al., 2018; Chen et al., 2018a; Chen and Konukoglu, 2018; Atlason et al., 2019) delineate anomalies by thresholding the resulting reconstruction errors, *i.e.*, the residual image between the input image *vs.* its reconstruction. Other methods (Martins et al., 2019a; Tang et al., 2019) train a one-class classifier from latent features to classify if an image (or region of interest) has some anomaly (Tang et al., 2019; Martins et al., 2019c). Although these methods can detect extensive lesions in MR-T2

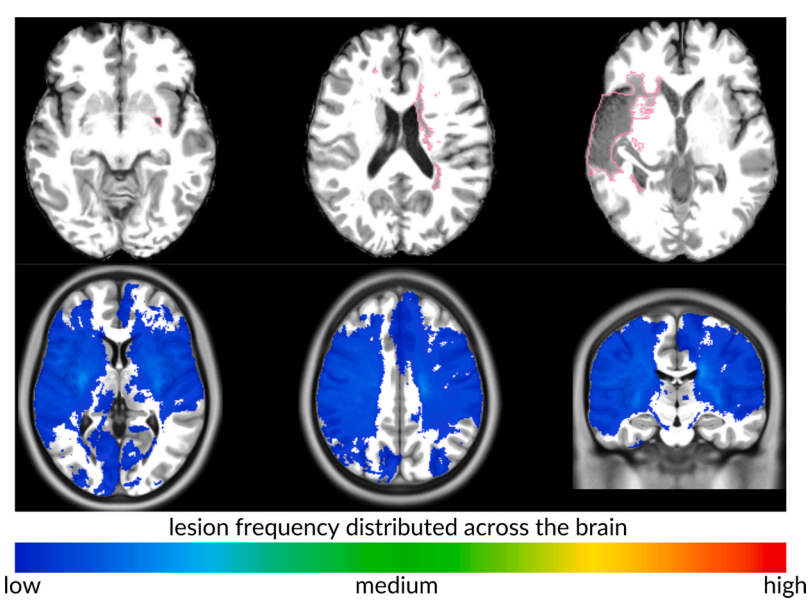


Fig. 1. The different appearance of brain anomalies. **Top:** axial slices of three stroke patients with lesions (gold-standard borders in pink) that significantly differ in location, shape, and size. **Bottom:** slices of a 3D heatmap show the location frequency of stroke lesions across the brain. Although caused by the same disease, the lesions are sparsely distributed in the brain resulting in low-concentrated regions. The 3D heatmap was built from aligned manual lesion segmentation of stroke patients from the ATLAS dataset (Liew et al., 2018) after registration to a standard template. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and CT (Sato et al., 2018; Chen et al., 2018a), they show inferior results in MR-T1 images and completely fail with small lesions, which are the most challenging cases.

Unsupervised approaches. All the above limitations of supervised methods motivate research on *unsupervised* anomaly detection approaches (Martins et al., 2019a; Sato et al., 2018; Baur et al., 2018; Chen et al., 2018a; Guo et al., 2015). These methods aim to learn a model from control images of healthy subjects *only* by encoding general knowledge or assumptions (*priors*) from healthy tissues. This model is next used to guide brain segmentation so that *outliers* who break such general priors are considered anomalies (Guo et al., 2015). However, as they do not use annotated samples, these methods may not accurately capture subtle differences between lesions and their surrounding healthy tissues. Shen et al. (2008) proved that the voxel-intensity-based segmentation and the spatial-location-based tissue distribution (based on a probabilistic atlas) in the lesions are inconsistent with those in healthy tissues. They use the conventional Fuzzy C-Mean algorithm and probabilistic maps from a template to quantify such inconsistencies, and then apply a threshold to obtain a binary lesion segmentation. Juan-Albarracín et al. (2015) propose a more complex method that uses four different image modalities of MRI for a given patient to segment the brain tumors of the BraTS dataset (Menze et al., 2014). All image voxels are grouped in a few clusters classified as normal or outlier, based on probabilistic maps from a template. However, this method is not practical in clinical routine because it requires many image modalities and only detects large lesions, as present, *e.g.*, in BraTS.

Since many neurological diseases are associated with abnormal brain asymmetries (Wang et al., 2001), a method called Supervoxel-based Abnormal Asymmetry Detection (SAAD) (Martins et al., 2019a) was recently proposed to detect abnormal asymmetries in MR brain images. SAAD registers all images to the same symmetric template, then computes asymmetries between the two hemispheres by using their mid-sagittal plane (MSP) as reference. Next, a new supervoxel segmentation method, named SymmISF, is used to extract pairs of symmetric supervoxels from the left and right hemispheres for each test image, guided by their asymmetries. Each pair generates a one-class classifier trained on control images to find supervoxels with abnormal asymmetries on the test image. SAAD was further extended to detect abnormal asymmetries in the own native image space of each test image (Martins and Telea, 2019). SAAD yields higher anomaly detection accuracy than deep generative neural networks (Baur et al., 2018; Chen et al., 2018a). However, as outlined in Section 1, the effect of the supervoxel segmentation quality and the use of a one-class classifier on the SAAD's detection accuracy are unknown.

3. Description of SAAD

We next describe the SAAD method (see also Fig. 2) and the supervoxel segmentation methods that we will use in conjunction to it. The method consists of four steps: 3D image preprocessing, asymmetry computation, supervoxel segmentation, and classification, described next. SAAD can detect abnormal asymmetries in any stage of a brain disorder, but it should preferably be used in clinical trials of initial stages to help early diagnosis and timely treatment.

3.1. 3D image preprocessing

Automated analysis of MR images is very challenging due to inherent image acquisition problems such as noise, intensity heterogeneity, and variability of the intensity range and contrast. To alleviate these and make images more similar to each other, we use typical preprocessing steps known in the literature (Akkus et al., 2017; Martins et al., 2019a; Manjón and Coupé, 2016; Juan-Albarracín et al., 2015), as shown in Fig. 3.

We preprocess the training control image set and the test image (Steps 1 and 4 in Fig. 2, respectively) by first performing noise reduction by median filtering, followed by mid-sagittal plane (MSP) alignment, and bias field correction by N4 (Tustison et al., 2010). Since voxels from irrelevant tissues/organs for the addressed problem (*e.g.*, neck and bones) can negatively impact the image registration and intensity normalization, we use the probabilistic atlas-based method AdaPro (Martins et al., 2019b) for skull stripping (Fig. 3b).

To attenuate differences in brightness and contrast among images, we first apply a histogram matching between the segmented images and the template (inside its predefined brain segmentation mask). This operation only considers voxels inside the brain (Fig. 3d). We then perform deformable registration to place all images in the coordinate space of the ICBM 2009c Nonlinear Symmetric template (Fonov et al., 2009), whose hemisphere masks and MSP are well defined. This is a popular and widely used template in the literature constructed by averaging 152 brain 3D MT-T1 images from healthy subjects aged 18–43 years. This averaging process relies on several image-preprocessing operators, such as intensity normalization and nonlinear registration. We refer to Fonov et al. (2009) for more details regarding the template construction. All registrations are performed by Elastix (Klein et al., 2010).[1]

Finally, we perform another histogram matching between the registered images and the template, and use the brain segmentation mask from the template and its MSP to separate the left and right brain hemispheres in each image for further asymmetry computation (Fig. 3e).

3.2. Asymmetry computation

Let X be the set of registered training 3D images (output of Step 1) and I the test 3D image after preprocessing (output of Step 4). We obtain the set of asymmetry maps A_X for all X by computing the voxel-wise absolute differences between left and right hemispheres concerning the template's MSP (Step 2).

Next, we create a *normal asymmetry map* $\overline{A_X}$ (Step 3) by averaging the absolute difference values of A_X (Fig. 4a). We use this map to reduce the detection of false-positive asymmetries in I in commonly asymmetric brain regions (*e.g.*, cortex), as detailed next in Section 5. Finally, we compute voxel-wise absolute differences between the hemispheres for I (Fig. 4b–c) and then subtract $\overline{A_X}$ from them. Resulting positive values form a final asymmetry map A_I (Fig. 4d) for the test image I (Step 5).

3.3. Symmetric supervoxel segmentation

Directly comparing the flipped, segmented, and registered hemispheres is not helpful as it will not tell us where *small-scale* asymmetries occur (Martins and Telea, 2019). At the other extreme, comparing every voxel pair in these hemispheres is risky, since individual voxels contain too little information to capture asymmetries. These difficulties motivate the use of *supervoxels* as the unit of comparison (Step 6).

An ideal supervoxel segmentation should create precisely one supervoxel per anomaly. This is, of course, highly unlikely to succeed, given the high variability of size, shape, and position of anomalies (see again Fig. 1). At any rate, too small supervoxels should be avoided as they *oversegment* larger anomalies and thus cannot capture their essence, and also will confuse the end users when visually exploring the results. Too large supervoxels, in contrast, should be avoided as they cannot precisely delineate small-scale anomalies from the background (*undersegmentation*).

SAAD uses a method called SymmISF (Martins et al., 2019a) to extract symmetrical supervoxels from left and right brain hemispheres simultaneously. SymmISF is based on the recent Iterative Spanning Forest (ISF) framework (Vargas-Muñoz et al., 2019) for superpixel

[1] We used the *par0000* files available at http://elastix.bigr.nl/wiki/index.php/Parameter_file_database.

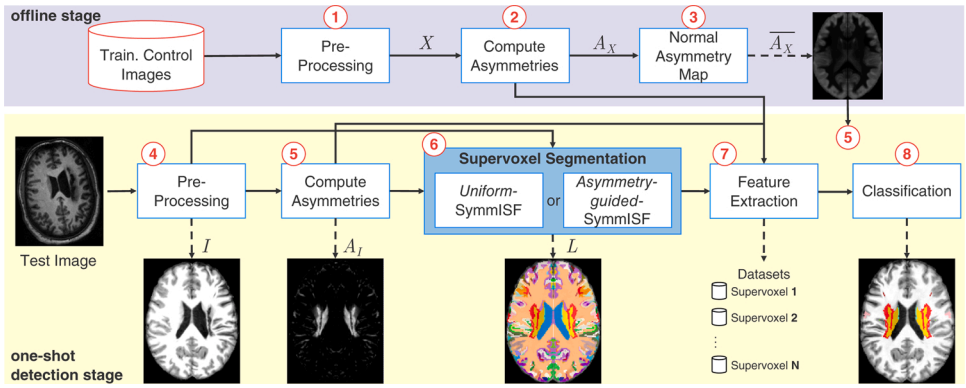


Fig. 2. Extended pipeline of SAAD with two possible symmetric supervoxel segmentations. Steps 1 to 3 (purple part) are performed offline. Steps 4 to 8 (yellow part) are computed for each test image (detection stage). The output images from Steps 3, 5, 6, and 8 are visualized as a symmetrical image. However, the method can consider just one hemisphere. Figure based on Martins et al. (2019a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

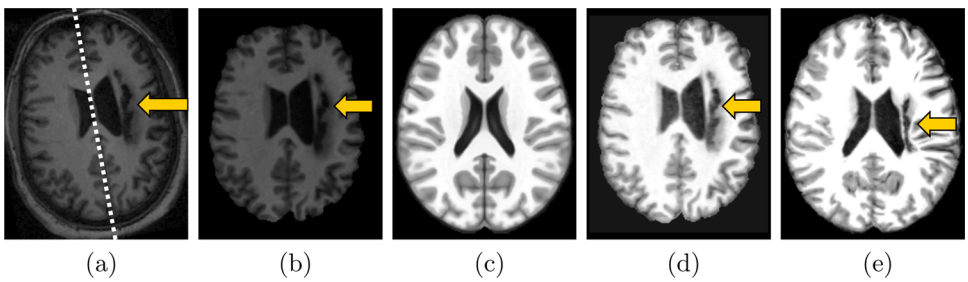


Fig. 3. 3D image preprocessing steps. **(a)** Axial slice of a raw test 3D image. The dashed line shows its mid-sagittal plane (MSP), and the arrow indicates a lesion. **(b)** Test image after noise filtering, MSP alignment, bias field correction, and brain segmentation. **(c)** Axial slice of the symmetric brain template (reference). **(d)** Histogram matching between (b) and the template (intensity normalization). **(e)** Final preprocessed image after deformable registration and histogram matching with the template. Figure referenced from Martins et al. (2020).

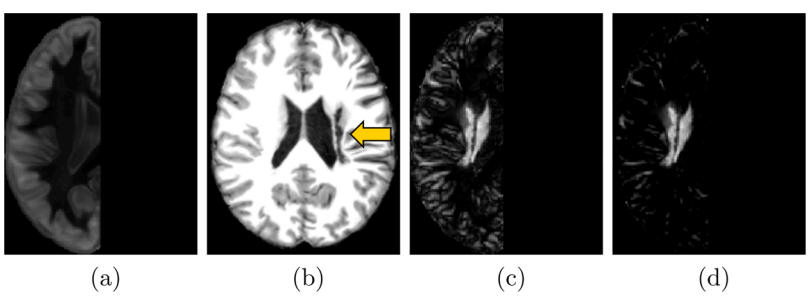


Fig. 4. Asymmetry computation on a standard image space. **(a)** Axial slice of the normal asymmetry map for healthy subjects. **(b)** Axial slice of a 3D test stroke image after preprocessing and deformable registration on a symmetric template. **(c)** Asymmetries of (b) by computing voxel-wise absolute differences between the hemispheres with respect to its MSP. **(d)** Final attenuated asymmetries: positive values of the subtraction between (c) and (a).

segmentation and has three steps: (i) initial seed estimation; multiple iterations of (ii) connected supervoxel delineation; and (iii) seed recomputation to improve delineation (Fig. 5), as follows.

Initial seed estimation is a crucial step for the success of ISF. The adopted strategy for that, however, depends on the target problem, which, in turn, may have specific constraints. For the problem of this work, the simplest approach to find initial seeds is to select N seeds uniformly distributed in the right hemisphere defined by a segmentation mask for the template. We call this strategy Uniform SymmISF next. However, there are no guarantees this strategy will place *at least* one seed within each asymmetric anomaly, so this can easily lead to undersegmentation.

The initial seed estimation strategy proposed by SAAD, called next Asymmetry-guided SymmISF, is guided by the hemispheric asymmetries of the image when selecting one seed per local maximum in A_I (see the asymmetry-guided seeds in Fig. 5). It computes the local maxima of the foreground of a binarized A_I at $\gamma \times \tau$, where τ is Otsu's threshold (Otsu, 1979). The higher the factor γ is, the lower is the number of asymmetric components in the binarized A_I. This seed-set is next extended with a fixed number (*e.g.*, 100) of seeds by uniform grid sampling the low-asymmetry regions of the binarized image. A detailed comparison of Uniform SymmISF with Asymmetry-guided SymmISF is presented next in Section 5.1.

As the cortex is typically very asymmetric, we can still remove seeds placed very close to the hemisphere borders to reduce the number of false positives in such regions and also to weight the normal asymmetry map to attenuate other asymmetries further. Both strategies are evaluated and discussed in detail in Section 5.2.

By stacking the right hemisphere with the left hemisphere — flipped using the MSP — as the input 2-band volume (Fig. 5), SymmISF applies ISF only inside the right hemisphere from the initial seeds. ISF relies on a cost function controlled by two parameters: α and β. This yields a label map in which each supervoxel is given a distinct label. Finally, SymmISF flips these supervoxels to obtain the symmetrical supervoxels in the left hemisphere, which yields the final label map L (output of Step 6). Note that one can proceed conversely, *i.e.*, apply SymmISF on the left hemisphere, and map the result to the right hemisphere.

3.4. Feature extraction and classification

SAAD presents a novel approach for outlier detection — here instantiated for abnormal asymmetry detection — that designs a set of

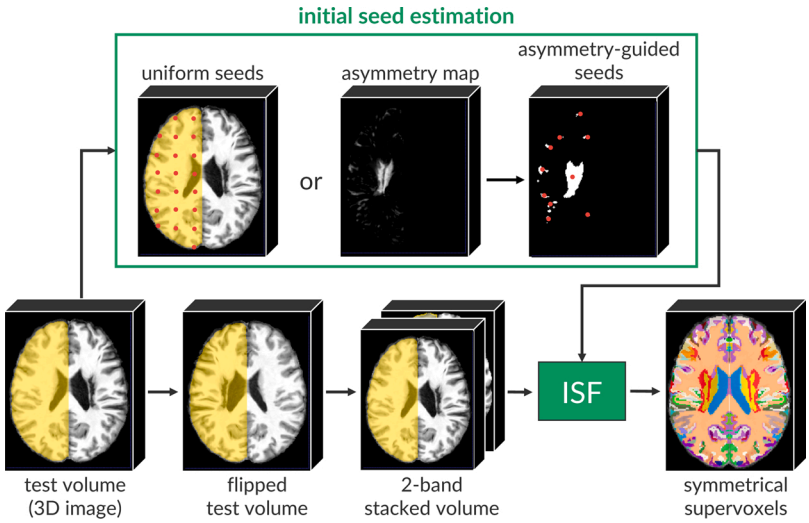


Fig. 5. The pipeline of SymmISF with two possible initial seed estimation strategies (red points represent seeds). The method flips the input test 3D image (volume) using its MSP and builds a 2-band volume by stacking both volumes. Then, the ISF framework (Vargas-Muñoz et al., 2019) estimates supervoxels inside the hemisphere mask from the initial seeds. The resulting label map is flipped to form the final label map with pairs of symmetrical supervoxels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

specialized one-class classifiers (OCCs) specific for each test 3D image, as shown in Fig. 6. For each 3D test image, each pair of symmetrical supervoxels is used to create a *specialized* OCC using as feature vector the *normalized histogram* of the asymmetry values inside the pair (Step 7). Classifiers are trained from *control images* only, thus locally modeling normal asymmetries for the entire hemispheres. Finally, SAAD uses the trained OCCs to find supervoxels with abnormal asymmetries in the corresponding testing image (Step 8). Fig. 7 illustrates the supervoxel classification.

By default, SAAD yields *pairs* of symmetric supervoxels corresponding to the detected abnormal asymmetries. This output is useful for subsequent visual analysis as an expert can compare such regions in both hemispheres as well as their computed asymmetries. To output only the supervoxel that covers the detected asymmetric anomaly, one may simply compute the similarity from the test image with the template inside each supervoxel of the pair. The less similar supervoxel contains the anomaly.

When *dynamically* designing specialized one-class per-supervoxel classifiers for each test image, SAAD implicitly considers the *position* of the supervoxels in the hemispheres when deciding upon their asymmetry. The central premise for this is that a *single* global classifier cannot separate normal and anomalous tissues by only using texture features. Experimental results concerning this hypothesis are presented in Section 5.3.

Even though the proposed classification scheme demands a higher processing time compared to using a single global classifier trained offline, this time is not too high (≈2 min) and still feasible for clinical purposes as SAAD relies on a simple and fast feature extraction (histogram) and the one-class linear Support Vector Machine (Manevitz and Yousef, 2001). Section 5 presents more details.

4. Experiments

This section describes the MR-T1 image datasets, baselines, and the evaluation protocol considered for the experiments. All computations were performed on the same Intel Core i7-7700 CPU 3.60 GHz with 64 GB of RAM.

4.1. Datasets

To answer our research questions (see Section 1), we need datasets with *volumetric* MR-T1 brain images (i) from *healthy* subjects (for SAAD training), and (ii) with hemispheric *asymmetric* lesions of different sizes (especially small ones) and their segmentation masks. For this, we first considered the CamCan dataset (Taylor et al., 2017), which has 653 MR-T1 images of 3T from healthy men and women between 18 and 88 years. For each 3D MR-T1 image, CamCan also has a corresponding 3D MR-T2 image, which we do not use in this work.

To our knowledge, CamCan is the largest public dataset with 3D images of healthy subjects acquired from different scanners. We visually inspected all MR-T1 images and removed images with bad acquisition or artifacts, yielding 524 images.[2]

Public datasets with different brain lesions exist. However, some only provide a subset of 2D slices for each image or interpolate slices to build a volume (*e.g.*, BraTS (Menze et al., 2014)); others provide 3D images with only very symmetric lesions (*e.g.*, MSSEG (Commowick et al., 2018)). Given these limitations, we settled on using the Anatomical Tracings of Lesions After Stroke (ATLAS) public dataset release 1.2 (Liew et al., 2018) in our experiments.

ATLAS is a challenging dataset with a large variety of manually annotated lesions and images of individuals after stroke acquired from different scanners. It contains lesions ranging from very small to large, located in several parts of the brain (see Fig. 1 for examples). All images have a mask with the primary stroke region. Some images also have additional masks with other stroke lesions.

ATLAS provides lesions from different phases/stages *after* stroke onset. As outlined in Section 3, although SAAD can detect abnormal asymmetries in *any* of these stages, it should preferably be used in the initial phase during clinical use to help early diagnosis and timely treatment — especially for stroke whose initial stage is aggressive. Other brain disorders, such as some types of cancer, can be identified in even earlier stages when such disorders are not still too aggressive as stroke. Thus, SAAD is even more helpful for these cases when providing early abnormal brain asymmetry detection. Given the absence of labeled masks for different anomalies, this evaluation only considers stroke lesions.

Since SAAD is designed to detect *hemispheric* abnormal asymmetries and the considered training images have a 3T field strength, we selected all 3T images from ATLAS which contain only lesions in the hemispheres (total of 229 images). All images were registered into the coordinate space of ICBM 2009c Nonlinear Symmetric template (Fonov et al., 2009) and preprocessed as outlined in Section 3.1.

4.2. Evaluation protocol

Baselines: In the absence of details, available tools, and trained

[2] A list with the selected images can be found on https://github.com/lidsunicamp/CMIG20_BrainAsymmetryDetection.

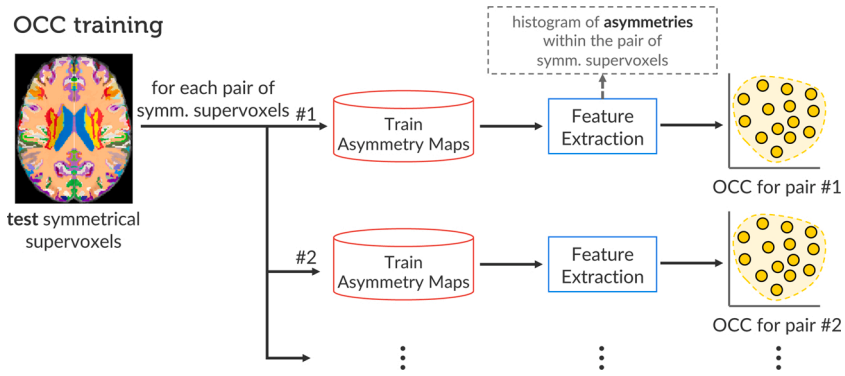


Fig. 6. One-class classifier (OCC) training for abnormal asymmetry detection. For each pair of symmetric supervoxels from a given test 3D image, SAAD trains an OCC from the training normal asymmetry maps previously computed.)

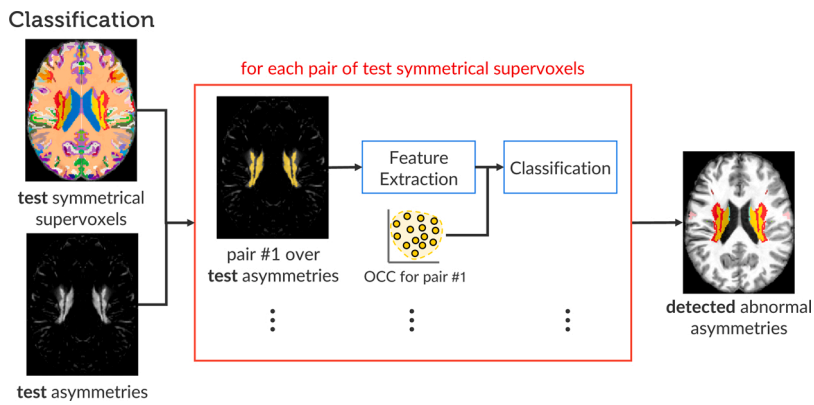


Fig. 7. Abnormal asymmetry detection of a test 3D image by supervoxel classification. For each pair of symmetrical supervoxels, SAAD uses the corresponding one-class classifier to classify the asymmetries inside the pair.

models for automatic anomaly detection, we compared SAAD against the convolutional-autoencoder-based approach (CAE) from Chen et al. (2018a), which is, as far as we know, the current state-of-the-art *unsupervised* method for the ATLAS dataset.

We considered the 2D axial slices of all preprocessed training images to train CAE, which has the following architecture: three 2D convolutional layers with 16, 8, and 8 filters of patch size 3×3, respectively, followed by *ReLU* activation and 2D max-pooling in the *encoder*, and the corresponding operations in the *decoder*. The nadam gradient optimizer minimized the mean squared error between reconstructed and expected 2D axial slices during training. The method detects anomalies by thresholding the resulting residual image of between the input image *vs.* its reconstruction to obtain a binary segmentation, similarly to Baur et al. (2018) and Chen et al. (2018a). We followed Baur et al. (2018) and selected three thresholds as the 85th, 90th, and 95th percentile from the histogram of reconstruction errors on the considered training set, resulting in the brightness of 143, 194, and 282, respectively. For simplicity, we call these three versions of the method as CAE-85, CAE-90, and CAE-95, respectively, based on the chosen percentiles.

Quality metrics: Although SAAD detects abnormal asymmetries regardless of the type of anomalies, we can compute quantitative scores only over those lesions that are labeled in ATLAS, which are a subset of what SAAD can detect. For these lesions, we first computed the *detection rate* based on at least 15% overlap between lesions labeled in ATLAS with detected volumes of interest (VOIs) with abnormal asymmetries (Tables 1–3, row 1), as detected by SAAD (supervoxels) and CAE (segmented regions). We then computed the *true positive rate* (recall) that measures the percentage of lesion voxels correctly classified as abnormal (Tables 1–3, row 2). Although our focus is on *detecting* abnormal asymmetries, we measured the *Dice score* between lesions and the detected VOIs to check SAAD's potential as a *segmentation method* (Tables 1–3, row 3). However, observe that truly abnormal asymmetries detected by SAAD that are not annotated as lesions in the ground-truth masks will be incorrectly considered as false positive and, thus, underestimating the Dice score. We then also measured the Dice score by considering only supervoxels overlapped with the annotated lesions (Tables 1–3, row 4).

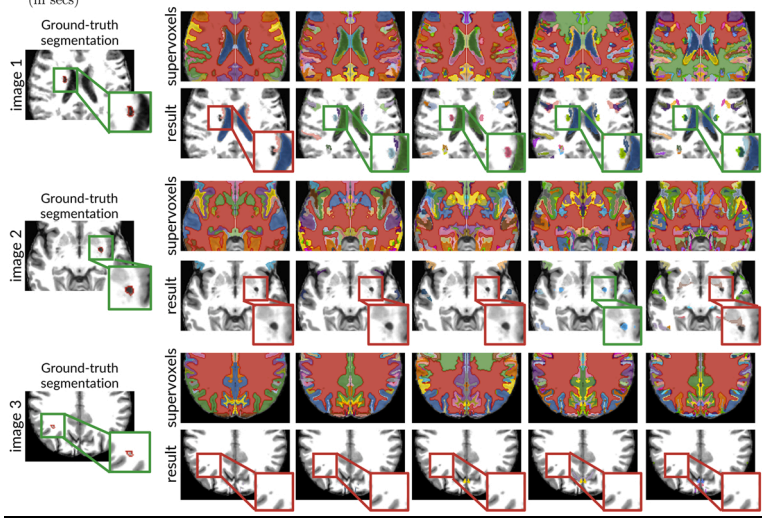
Highly accurate detection methods are useful only if their false positive count is quite low. Otherwise, one needs to manually inspect the many positives to validate them, which is very costly. To gauge this, we provided false-positive (FP) scores in terms of both voxels and supervoxels with respect to the ground-truth stroke lesions of ATLAS. Hence, some regions with true abnormal asymmetries but with no labeled masks in ATLAS are considered FP (*e.g.*, see the deformed ventricles of the third image of Fig. 1). This is the best we can do in the absence of labeled masks for all kinds of abnormalities in this dataset.

To evaluate the detection quality, we propose a set of fine-to-coarse metrics, as follows. At the finest level, we first compute the mean number of FP voxels, *i.e.*, incorrectly classified as abnormal (Tables 1–3, row 5). We normalized this count regarding all classified voxels (Tables 1–3, row 6), *i.e.*, the total number of voxels inside the right hemisphere for SAAD, and both hemispheres for CAE. At the next level, we estimated FP supervoxels as those whose voxels overlap less than 15% with ground-truth lesion voxels. We computed the mean number of FP supervoxels and their proportions concerning the total number of supervoxels (Tables 1–3, rows 7 and 8). The first metric gives us an estimation of the visual-inspection user effort. The second metric checks how imprecise is the detection regarding the total number of regions that the user has to visually analyze.

When visually analyzing FP supervoxels, it is harder to check many

Table 1
Experimental results for SAAD with Uniform SymmISF for different numbers of seeds. **Top part:** higher values mean better accuracies. **Bottom part:** lower values mean better accuracies. Each result contains a box (inset) surrounding the lesion whose border color indicates if the lesion was detected (green) or missed (red). The abbreviation *k* denotes thousands.

		Uniform SymmISF				
		N=100	N=250	N=400	N=550	N=700
1	Detection rate	0.389 ± 0.058	0.62 ± 0.052	0.738 ± 0.041	0.808 ± 0.033	**0.86 ± 0.028**
2	True positive rate (mean recall)	0.159 ± 0.024	0.293 ± 0.025	0.364 ± 0.022	0.409 ± 0.017	**0.447 ± 0.018**
3	Dice	0.12 ± 0.022	**0.13 ± 0.02**	0.127 ± 0.02	0.126 ± 0.019	0.123 ± 0.018
4	Dice (only overlapped supervoxels with lesions)	0.152 ± 0.025	0.245 ± 0.025	0.298 ± 0.02	0.338 ± 0.016	**0.372 ± 0.015**
5	# FP voxels	**14k ± 1.36k**	26k ± 1.69k	32k ± 1.71k	36k ± 1.75k	40k ± 1.77k
6	FP voxel rate	**0.018 ± 0.002**	0.033 ± 0.002	0.041 ± 0.002	0.045 ± 0.002	0.05 ± 0.002
7	# FP supervoxels	**8.83 ± 0.27**	30.54 ± 0.84	49.97 ± 1.20	67.12 ± 1.86	90.91 ± 2.19
8	FP supervoxel rate	**0.103 ± 0.003**	0.122 ± 0.003	0.128 ± 0.003	0.128 ± 0.004	0.13 ± 0.003
9	# FP connected supervoxels	**6.39 ± 0.17**	18.40 ± 0.37	26.11 ± 0.47	32.67 ± 0.68	41.03 ± 0.76
10	FP connected supervoxel rate	**0.077 ± 0.002**	0.078 ± 0.002	0.071 ± 0.001	0.067 ± 0.001	0.063 ± 0.001
11	Mean processing time (in secs)	39.35 ± 0.98	59.08 ± 0.90	83.68 ± 1.00	111.77 ± 1.49	162.24 ± 12.59



disconnected supervoxels spread across the brain (*e.g.*, Fig. 8a) than a few connected ones (see, *e.g.*, the ventricle area in Fig. 8b). Hence, at the coarsest level, we gauged visual analysis user-effort by evaluating the two metrics outlined above on the level of connected FP supervoxel components (Tables 1–3, rows 9 and 10).

Finally, we also computed the mean processing times of each method (Tables 1–3, row 11) for preprocessed images, thus excluding the mean time of the preprocessing step (Section 3.1), which is 90 s on average.

5. Results

We next discuss our results from the perspective of our key questions, stated in Section 1.

5.1. Impact of supervoxel segmentation quality on abnormal asymmetry detection

To check if the supervoxel segmentation influences the abnormal asymmetry detection, we used two variants of the SymmISF method to extract pairs of symmetric supervoxels (one for each hemisphere), as follows.

Uniform SymmISF. This method chooses N initial seeds *uniformly distributed* in a hemisphere, with N defined by the user (Section 3.3). While simple, this strategy does not guarantee to place *at least* one seed within each asymmetric anomaly, especially when N is small. In turn, this leads to undersegmentation — the missed lesions will be assimilated to the background. Conversely, when N is too large, this easily leads to the oversegmentation of larger lesions into many supervoxels that have too little individual information to capture asymmetries.

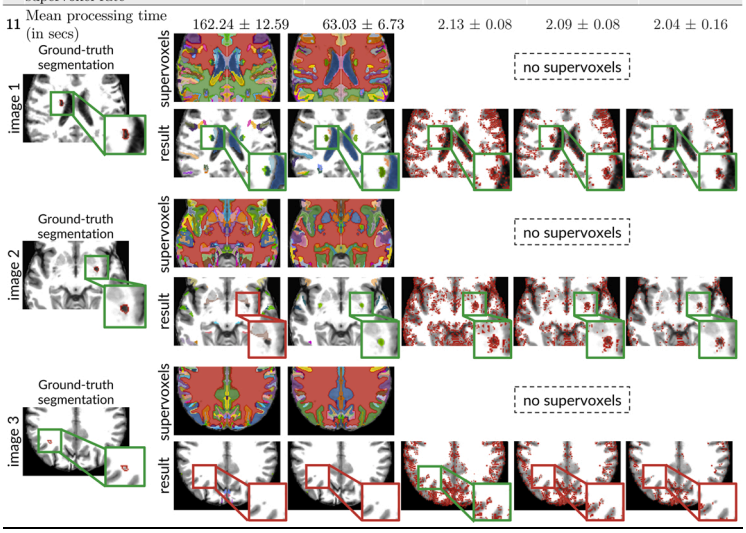
Asymmetry-guided SymmISF. To better fit supervoxels with asymmetric anomalies of various morphologies, this strategy first seeds the highest-asymmetry-value brain regions (where anomalies are more likely to occur) and then seeds the remaining, more symmetric, areas with a fixed number of extra seeds (Section 3.3). Since asymmetries vary for each image, the final number of supervoxels is dynamically obtained. For the experiments, we fixed 100 extra seeds uniformly distributed on low asymmetric regions of the images.

We quantitatively compare the above two initial seeding strategies using 5-fold cross-validation on ATLAS, considering one subset for *validation* (46 images) and the remaining four subsets for *testing* (183 images) in each fold. For this initial experiment, we use the original parameters for SAAD reported in Martins et al. (2019a), *i.e.*, $\alpha = 0.08$, $\beta = 3.0$, asymmetry histograms of 128 bins, and $\nu = 0.1$ for the linear one-class SVM. For Uniform SymmISF, we consider five different numbers of seeds N: {100, 250, 400, 550, 700}. For Asymmetry-guided SymmISF, we use $N = 100$ seeds in symmetric regions.

Table 1 shows the mean results of SAAD with Uniform SymmISF for the primary stroke lesions of ATLAS considering all five folds, and selected visual results. As expected, Uniform SymmISF presents poor detection results for low N values since anomalies are covered by large

Table 2
Quantitative and qualitative comparison between the two versions of SAAD and CAE (with different thresholds) for the ATLAS dataset. **Top part:** higher values mean better accuracies. **Bottom part:** lower values mean better accuracies. Each result contains a box (inset) surrounding the lesion whose border color indicates if the lesion was detected (green) or missed (red). The abbreviation *k* denotes thousands.

		Uniform SymmISF N=700	Asymmetry-guided SymmISF	CAE-85	CAE-90	CAE-95
1	Detection rate	0.86 ± 0.028	0.851 ± 0.016	**0.995 ± 0.002**	0.943 ± 0.018	0.55 ± 0.03
2	True positive rate (mean recall)	**0.447 ± 0.018**	0.436 ± 0.009	0.439 ± 0.01	0.333 ± 0.011	0.199 ± 0.011
3	Dice	0.123 ± 0.018	**0.132 ± 0.02**	0.018 ± 0.003	0.017 ± 0.003	0.016 ± 0.002
4	Dice (only overlapped supervoxels with lesions)	0.372 ± 0.015	**0.401 ± 0.007**	method does not use supervoxels		
5	# FP voxels	40k ± 1.77k	**28k ± 0.55k**	644k ± 2.55k	428k ± 2.1k	206k ± 3.68k
6	FP voxel rate	0.05 ± 0.002	**0.035 ± 0.001**	0.4 ± 0.002	0.267 ± 0.001	0.129 ± 0.002
7	# FP supervoxels	90.91 ± 2.19	**58.21 ± 1.83**	method does not use supervoxels		
8	FP supervoxel rate	**0.13 ± 0.003**	0.194 ± 0.004			
9	# FP connected supervoxels	41.03 ± 0.76	**29.81 ± 0.88**			
10	FP connected supervoxel rate	**0.063 ± 0.001**	0.111 ± 0.002			
11	Mean processing time (in secs)	162.24 ± 12.59	63.03 ± 6.73	2.13 ± 0.08	2.09 ± 0.08	2.04 ± 0.16



supervoxels that mix lesion and background voxels (see images in Table 1). As N increases, the chance of placing at least one seed inside each lesion is higher, even for smaller lesions, which leads to better results: We see how the detection rates, mean recall, and Dice (with only overlapped supervoxels with lesions) monotonically increase with N in Table 1 (rows 1, 2, and 4). Yet, there is no guarantee that increasing N yields increasingly-better fitting supervoxels to lesions. This is visible in the results for image 2 (insets) in Table 1, where we see that a small lesion was missed for $N = 700$ but found for $N = 550$. Moreover, the number of FP voxels and supervoxels also increases as N increases — compare rows 5, 7, and 9 of Table 1. This results in considerably high FP rates for large N values and explains the similar global Dice scores from all methods (row 3). Hence, visual inspection becomes difficult, even when the detection rate is high (compare row 1 and rows 5–10 for $N = \{550, 700\}$). Also, the more supervoxels we extract, the longer is the processing time, as shown in Table 1, row 11.

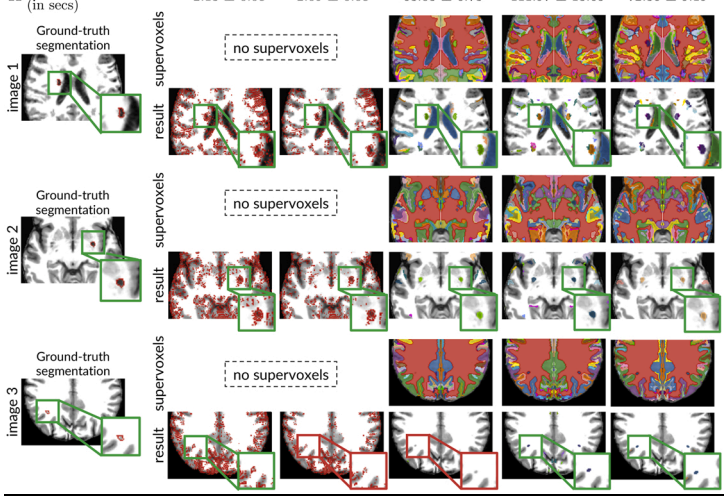
We next compare the Uniform and Asymmetry-guided SymmISF versions for SAAD against three versions of CAE, as presented in Table 2. For simplicity, we consider only the version of Uniform-SymmISF with the best detection accuracy. Also, note that only SAAD reports false-positive supervoxel-based metrics.

CAE-85 and CAE-90 present considerably higher detection scores, 0.995 and 0.943, respectively, than the two versions of SAAD: Uniform SymmISF (0.86) and Asymmetry-guided SymmISF (0.851). However, these impressive results are misleading as CAE reports drastically (about 20x) more false-positive voxels than SAAD — compare rows 5 and 6 in Table 2. For instance, although CAE-85 almost detects all lesions, it misclassifies 40% of the hemispheres as abnormal, which is far from being reasonable and hinders the visual analysis (we expect just a small portion of the brain, *e.g.*, 1%). These high FP rates explain the poor Dice scores for CAE in Table 2, which in turn are compatible with the ones reported in Chen et al. (2018a). Additionally, CAE yields very noisy detected regions, especially in regions with transitions between white and gray matter (*e.g.*, the cortex), which also hinders the subsequent visual inspection (see the results for the considered images in Table 2). Even though its FP voxels decrease as higher thresholds are considered, its detection score can be hugely impacted; for example, the threshold at the 95th percentile approximately halves both the detection score and FP voxels rates compared with the results for the 90th percentile in Table 2. CAE is speedy (running time about 2s per image) and might present better results by using a considerable large training set and/or some additional post-processing, but this is not considered in Chen et al. (2018a), Baur et al. (2018). CAE presents better results for other medical imaging modalities, such as CT and T2 (Chen et al., 2018a; Baur et al., 2018).

Asymmetry-guided SymmISF has a slightly worse detection rate (0.851) compared to Uniform SymmISF (0.86). Although it is also able to find small abnormal asymmetries (Table 2, images 1–2), it fails to detect very subtle and/or tiny asymmetries (Table 2, image 3). Also, this seeding strategy has lowest number of FP (connected) supervoxels and FP voxel scores. However, the expert still has to unnecessarily analyze about 29 FP connected supervoxels per image, which may take a

Table 3
Quantitative and qualitative comparison between SAAD after parameter optimization, and the baselines for the ATLAS dataset. **Top part:** higher values mean better accuracies. **Bottom part:** lower values mean better accuracies. Each result contains a box (inset) surrounding the lesion whose border color indicates if the lesion was detected (green) or missed (red). The abbreviation *k* denotes thousands.

		CAE-85	CAE-90	Asymmetry-guided SymmISF	Optimized-SymmISF	Optimized-SymmISF with FPA
1	Detection rate	**0.995 ± 0.002**	0.943 ± 0.018	0.851 ± 0.016	0.939 ± 0.008	0.862 ± 0.013
2	True positive rate (mean recall)	0.439 ± 0.01	0.333 ± 0.011	0.436 ± 0.009	**0.489 ± 0.006**	0.451 ± 0.006
3	Dice	0.018 ± 0.003	0.017 ± 0.003	0.132 ± 0.02	0.123 ± 0.018	**0.19 ± 0.018**
4	Dice (only overlapped supervoxels with lesions)	method does not use supervoxels		0.401 ± 0.007	0.397 ± 0.012	**0.411 ± 0.01**
5	# FP voxels	644k ± 2.55k	428k ± 2.1k	28k ± 0.55k	39k ± 1.87k	**11k ± 1.17k**
6	FP voxel rate	0.4 ± 0.002	0.267 ± 0.001	0.035 ± 0.001	0.049 ± 0.002	**0.014 ± 0.002**
7	# FP supervoxels	method does not use supervoxels		58.21 ± 1.83	98.15 ± 1.53	**21.19 ± 0.87**
8	FP supervoxel rate			0.194 ± 0.004	0.176 ± 0.003	**0.065 ± 0.005**
9	# FP connected supervoxels			29.81 ± 0.88	54.59 ± 0.73	**15.98 ± 0.56**
10	FP connected supervoxel rate			0.111 ± 0.002	0.107 ± 0.001	**0.049 ± 0.004**
11	Mean processing time (in secs)	2.13 ± 0.08	2.09 ± 0.08	63.03 ± 6.73	111.97 ± 13.65	72.36 ± 9.19



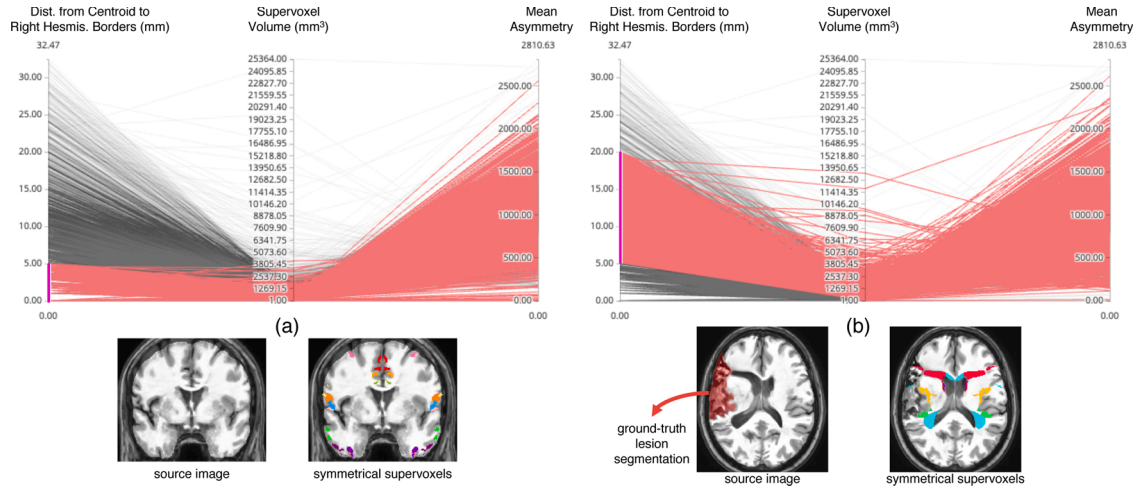


Fig. 8. Correlation between some characteristics of false-positive supervoxels. Each false positive supervoxel is a line in each plot which correlates the distance from its centroid to the right hemisphere's borders, its volume, and the mean asymmetry inside it. Red lines indicate supervoxels with distances **(a)** in [0, 5] mm and **(b)** in (5, 20] mm. An example of these corresponding supervoxels is shown below each plot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

considerable time. The next section details our strategy to improve SAAD with Asymmetry-guided SymmISF to yield higher detection rates and still attenuate FP scores.

5.2. Improving the end-to-end method

SAAD with Asymmetry-guided SymmISF is more suitable for our task

since each image's hemispheric asymmetries guide its supervoxel estimation. Moreover, as Table 2 shows, the asymmetry-guided seeding scales computationally better, being roughly twice as fast as uniform seeding for a comparable quality. Hence, we decided to improve Asymmetry-guided SymmISF by (i) optimizing its parameters by grid search aiming to increase detection accuracy; and (ii) proposing a false-positive-attenuation (FPA) strategy. We describe these optimizations next.

For parameter optimization, we considered the *validation* set of each fold (Section 4) and the following search space: $\alpha \in \{0.04, 0.06, 0.08, 0.1, 0.12\}$, $\beta \in \{1.0, 3.0, 5.0, 7.0, 9.0\}$, and $\gamma \in \{0.5, 1.0, 1.5, 2.0, 2.5\}$. As for cost function, we considered the *Intersection over Union* (IoU) metric that computes the overlap of a supervoxel with each lesion. Indeed, when this overlap is maximal, each lesion is accurately covered by precisely *one* supervoxel. The best parameters found by the grid search were $\alpha = 0.12$, $\beta = 5$, and $\gamma = 0.5$. Note that we used IoU as cost function, and not the metrics listed in Tables 1–2 since it generically looks at how supervoxels fit lesions, whereas those metrics gauge higher-level, more task-specific, concerns. Moreover, the IoU metric is continuous, making it usable for more refined parameter optimization, *e.g.*, using gradient descent. The metrics in Tables 1–2 are, in contrast, discrete, and do not allow such refined optimization.

We repeated the same experiment by considering the optimized parameters. Table 3 presents the results for this seeding strategy, called next Optimized-SymmISF. We replicated the results of CAE-85 and CAE-90, and Asymmetry-guided SymmISF in Table 3 to make the comparison easier. We see that Optimized-SymmISF has a higher detection rate (0.939) than Asymmetry-guided SymmISF (0.851), being slightly worse than CAE-90 (0.943). Optimized-SymmISF also presents the highest true positive rate (0.4889) among *all* compared methods. However, it still has high FP rates and has a considerable increase for the mean number of FP supervoxels and connected supervoxels (Table 3, rows 7 and 9) than Asymmetry-guided SymmISF.

To attenuate FPs, we first performed an analysis of their characteristics. Fig. 8 presents two parallel coordinate plots (PCPs) correlating the following three metrics on each FP supervoxel **s**: (i) distance d of centroid of **s** to the right hemisphere border; (ii) volume of **s**; and (iii) the mean asymmetry value inside **s**. Both PCPs are identical except by their highlighted examples (in red).

Fig. 8a highlights FP supervoxels close to the hemisphere's border ($d < 5$ mm), *i.e.*, in the cortex. These supervoxels are relatively small with high variability of mean asymmetries inside them and usually cover gyri and sulci (see the brain slices below the PCP), which are naturally asymmetric. Conversely, larger FP supervoxels are farther from the hemisphere border (Fig. 8b), although their mean asymmetries have high variability. By visually inspecting them, we can also find true abnormal regions deformed by the stroke lesions in the dataset (see the ventricles in the brain slices below the second PCP). Hence, it seems reasonable to reduce false- positive supervoxels in the cortex.

To do this, we propose a false-positive-attenuation (FPA) strategy that accentuates the normal asymmetry map (Section 3.2) by adding the standard-deviation asymmetries from the training set to it. As a result, the asymmetry map of the test image (output of Step 5 in Fig. 2) is more attenuated so that only highly asymmetric supervoxels will be detected as abnormal. Next, we remove the initial seeds found by Asymmetry-guided SymmISF whose distance to the hemisphere border is less or equal to 5 mm, as suggested in Fig. 8a. We repeated the same parameter optimization for SAAD using FPA, finding the optimal values $\alpha = 0.06$, $\beta = 5$, and $\gamma = 0.5$. Then, we repeated the full detection experiment for the optimal parameter method (called Optimized-SymmISF with FPA).

Table 3 (rightmost column) shows the results. Optimized-SymmISF with FPA has slightly better detection rate (0.862) and TP rate (0.451) to Asymmetry-guided SymmISF. Also, it can detect subtle and tiny asymmetric lesions (Table 3, image 3), which indeed are well-defined by its supervoxels. Although its detection rate is lower than Optimized-SymmISF, it attains the *lowest* FP rates from all considered methods (compare rows 5–10 in Table 3). This method yields, on average, only 1.40% of all voxels as FPs, and these cover only 4.9% of all connected supervoxels. Moreover, Optimized-SymmISF with FPA yields about from twice to three times less FP connected supervoxels for visual analysis than the other versions of SAAD, which decreases the user effort. Hence, Optimized-SymmISF with FPA has the best balance between high detection rates and low FP rates from all studied methods.

Next, we compared Optimized-SymmISF with FPA, our best method so far, with other related methods in the literature. Such methods are usually designed for the segmentation of, *e.g.*, organs or lesions. As we do not have access to implementations of these methods running on the same dataset as ours, except for CAE, we cannot compute all metrics shown in Table 3. The best we can do is to compare our method with these alternatives as a *segmentation* tool, using segmentation scores. However, note, again, that our method is designed primarily for *detection*, not segmentation.

Optimized-SymmISF with FPA yields the best Dice score (0.19) among all compared methods in the experiments. As outlined in Section 4.2, however, this score is underestimated as truly abnormal asymmetries detected by our method, which are not labeled as lesions in the ground-truth masks, are considered false-positive. When considering only symmetric supervoxels overlapped with the annotated lesions (Table 3, row 4)), such a Dice score leverages to 0.411. While still low, this score is not far from state-of-the-art results (Dice score 0.4867) on the ATLAS dataset from a *supervised* method based on U-Net (Qi et al., 2019). Interestingly, our method is noticeably superior to CAE, which is an *unsupervised* method (like ours), reporting very low Dice scores of 0.018 and 0.017 for thresholds at the 85th and 90th percentile, respectively. Our method reports drastically fewer FP voxels than CAE. Also, note that these compared methods yield their above-reported Dice scores by segmenting quite *large* lesions; in contrast, we focus on the more challenging problem of finding many *small* lesions (see, *e.g.*, image 3 in Table 3).

5.3. Per-supervoxel vs. global classifier design

We now investigate our second research question, *i.e.*, whether a per-supervoxel classifier design is indeed preferable to a global classifier. Suppose (hypothetically) that supervoxel segmentation is *completely irrelevant* for the final detection accuracy. Thus, the features we use (normalized histogram of absolute asymmetries for each symmetrical supervoxel) should be able to yield robust texture features for detection regardless of supervoxel quality. Hence, *only* a single classifier — not a (specialized) classifier per supervoxel for each test image — trained from texture features of training images should be enough to obtain similar results to those in Tables 1–3.

To test this hypothesis, we first chose a brain image from ATLAS (Fig. 9) with a large asymmetric stroke lesion, which is not as challenging to detect as a small one. If our hypothesis were correct, this lesion should be classified *easily* by the global classifier. If global classification failed, then the situation would be even worse for smaller, harder to detect, lesions. To investigate this further, we projected the texture feature vectors of all symmetric supervoxels extracted by Optimized-SymmISF with FPA using t-SNE (Maaten and Hinton, 2008) (Fig. 9). Here, each point represents a symmetric supervoxel, colored by its overlapping percentage with the ground-truth lesion. We see that there is no clear separation between the high-overlap supervoxels (warm-colored points) and healthy-tissue supervoxels (cool colors), even though the considered lesion is very well-defined by a single supervoxel (compare the brain images in Fig. 9). It has been shown that the visual separability of classes in a t-SNE projection is highly correlated with the ability of a classifier to separate classes in the original feature space (Rauber et al., 2017). Hence, since we do not find good visual separation, we conclude that a single classifier based only on texture features is insufficient to detect lesions, even large ones.

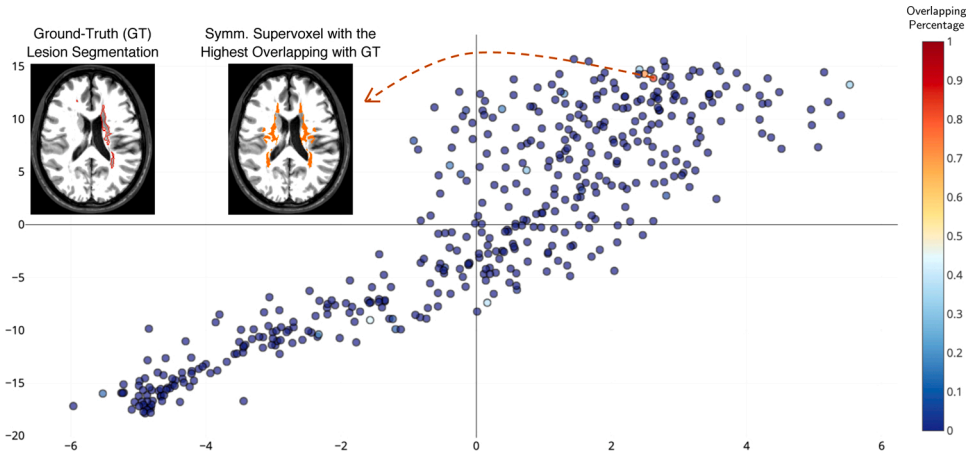


Fig. 9. t-SNE projection (perplexity of 50) from texture feature vectors (normalized histogram of absolute asymmetries) for the symmetric supervoxel extracted by Optimized-SymmISF with FPA for a given stroke image. The overlapping percentage between the ground-truth lesion segmentation and each supervoxel is color-encoded in the plot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusion

We presented a detailed investigation of the impact of supervoxel segmentation for unsupervised abnormal brain asymmetry detection. To this end, we chose a recent supervoxel-based approach (SAAD) that detects abnormal asymmetric lesions of a given target image by classifying extracting pairs of symmetric supervoxels by using a model (one-class classifier) trained for each pair, independently, from healthy brain anatomies only. Although SAAD presents high detection accuracies, there was, prior to our study, no evidence of the impact of supervoxels segmentation for the abnormal asymmetry detection, as well as the need to use a specialized one-class classifier for each pair of supervoxel instead of a global classifier.

We used SAAD to detect asymmetric stroke lesions on 3D MR-T1 brain images from a wide range of different symmetric supervoxels extracted by two different methods. Experimental results show that the quality of supervoxel segmentation *truly* impacts anomaly detection, especially for small anomalies. This analysis also helped us to find optimal parameter values and an improved seeding strategy that further improved quantitative results (*e.g.*, detection rate and false-positive rates) compared to the original SAAD method and a state-of-the-art unsupervised approach. Next, we showed that a single global classifier only based on texture features is not sufficient to detect even large anomalies, since their textures are similar to some healthy brain tissues. Putting together our experimental insights, we conclude that (1) a good fit of symmetrical supervoxels to lesions and (2) using a per-supervoxel classifier *are* beneficial design decisions for a proper detection of abnormal asymmetries.

For future work, we initially plan to improve brain asymmetry computation and investigate other feature extraction techniques for SAAD. We then intend to use additional visual analytics techniques to find challenging cases where the optimized SAAD method fails to detect complex small-scale lesions and improve seeding, asymmetry computation, feature extraction, or classification to yield better detection rates.

Authors' contributions

Conception: SBM, ACT, AXF.
Study design, methods used: SBM, ACT, AXF.
Acquisition of data: SBM.
Analysis and/or interpretation of data: SBM, ACT, AXF.
Writing the manuscript: SBM, ACT, AXF.
Critical revision of paper: SBM, ACT, AXF.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgment

The authors thank CNPq (303808/2018-7) and FAPESP (2014/12236-1) for the financial support, and NVIDIA for providing a graphics card.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Akkus, Z., et al., 2017. Deep learning for brain MRI segmentation: state of the art and future directions. J. Digit. Imaging 30, 449–459.
Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46, 726–738.
Aslani, S., Dayan, M., Murino, V., Sona, D., 2018. Deep 2D encoder-decoder convolutional neural network for multiple sclerosis lesion segmentation in brain MRI. Medical Image Computing and Computer-Assisted Intervention (MICCAI) 132–141.
Atlason, A., Love, H.E., Sigurdsson, S., Gudnason, V., Ellingsen, L.M., 2019. Unsupervised brain lesion segmentation from MRI using a convolutional autoencoder. SPIE Med. Imag., vol. 10949 372–378.
Baur, C., Wiestler, B., Albarqouni, S., Navab, N., 2018. Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images. arXiv preprint arXiv: 1804.04488.
Chen, X., Konukoglu, E., 2018. Unsupervised Detection of Lesions in Brain MRI Using Constrained Adversarial Auto-Encoders. arXiv preprint arXiv:1806.04972.
Chen, X., Pawlowski, N., Rajchl, M., Glocker, B., Konukoglu, E., 2018a. Deep Generative Models in the Real-World: An Open Challenge From Medical Imaging. arXiv preprint arXiv:1806.05452.
Chen, H., et al., 2018b. VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. Neuroimage 170, 446–455.
Commowick, O., et al., 2018. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Sci. Rep. 8, 13650.
Fonov, V.S., et al., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage 47, S102.
Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N., 2011. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. Neuroimage 57, 378–390.
Goetz, M., et al., 2014. Extremely randomized trees based brain tumor segmentation. Proc. of BRATS Challenge-MICCAI, 006-011.
González-Villà, S., et al., 2019. Brain structure segmentation in the presence of multiple sclerosis lesions. Neuroimage 22, 101709.
Guo, D., et al., 2015. Automated lesion detection on MRI scans using combined unsupervised and supervised methods. BMC Med. Imaging 15, 50.
Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2017. Brain tumor segmentation with deep neural networks. Med. Image Anal. 35, 18–31.

Juan-Albarracín, J., Fuster-Garcia, E., Manjón, J.V., Robles, M., Aparici, F., Martí-Bonmatí, L., García-Gómez, J.M., 2015. Automated glioblastoma segmentation based on a multiparametric structured unsupervised classification. PLOS ONE 10, e0125143.
Kamnitsas, K., et al., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.
Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J., 2010. elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29, 196–205.
Kooi, T., Litjens, G., Van Ginneken, B., Gubern-Mérida, A., Sánchez, C.I., Mann, R., den Heeten, A., Karssemeijer, N., 2017. Large scale deep learning for computer aided detection of mammographic lesions. Med. Image Anal. 35, 303–312.
Lötjönen, J.M.P., et al., 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage 49, 2352–2365.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436.
Liew, S.-L., et al., 2018. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Sci. Data 5, 180011.
Maaten, L.v.d., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.
Manevitz, L.M., Yousef, M., 2001. One-class SVMs for document classification. J. Mach. Learn. Res. 2, 139–154.
Manjón, J.V., Coupé, P., 2016. volBrain: an online MRI brain volumetry system. Front. Neuroinform. 10.
Martins, S.B., Telea, A.C., Falcão, A.X., 2019. Extending supervoxel-based abnormal brain asymmetry detection to the native image space. IEEE Eng in Med Bio Society (EMBC) 450–453.
Martins, S.B., Ruppert, G., Reis, F., Yasuda, C.L., Falcão, A.X., 2019a. A supervoxel-based approach for unsupervised abnormal asymmetry detection in MR images of the brain. Proc. IEEE ISBI 882–885.
Martins, S.B., Bragantini, J., Yasuda, C.L., Falcão, A.X., 2019b. An adaptive probabilistic atlas for anomalous brain segmentation in MR images. Med. Phys. 46, 4940–4950.
Martins, S.B., Benato, B.C., Silva, B.F., Yasuda, C.L., Falcão, A.X., 2019c. Modeling normal brain asymmetry in MR images applied to anomaly detection without segmentation and data annotation. SPIE Med. Imag. 109500C.
Martins, S.B., Falcão, A.X., Telea, A.C., 2020. BADRESC: brain anomaly detection based on registration errors and supervoxel classification. Biomedical Engineering Systems and Technologies: BIOIMAGING 74–81. Best student paper awards.
Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., 2014. The multimodal brain tumor image segmentation benchmark (brats). IEEE Trans. Med. Imaging 34, 1993–2024.
Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66.
Pinto, A., Pereira, S., Correia, H., Oliveira, J., Rasteiro, D.M., Silva, C.A., 2015. Brain tumour segmentation based on extremely randomized forest with high-level features. IEEE Eng in Med Bio Society (EMBC) 3037–3040.
Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M.P., Shyu, M.-L., Chen, S.-C., Iyengar, S.S., 2018. A survey on deep learning: algorithms, techniques, and applications. ACM Comput. Surv. 51, 92.
Qi, K., et al., 2019. X-net: brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies. Medical Image Computing and Computer-Assisted Intervention (MICCAI) 247–255.
Rauber, P.E., et al., 2017. Projections as visual aids for classification system design. Inf. Vis.
Sato, D., et al., 2018. A primitive study on unsupervised anomaly detection with an autoencoder in emergency head ct volumes. SPIE Med. Imag. 105751P.
Shen, S., Szameitat, A.J., Sterr, A., 2008. Detection of infarct lesions from single mri modality using inconsistency between voxel intensity and spatial location-a 3D automatic approach. IEEE Trans. Inf. Technol. B 12, 532–540.
Soltaninejad, M., et al., 2017. Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. Int. J. Comput. Assist. Radiol. 12, 183–203.
Tang, Y., Tang, Y., Han, M., Xiao, J., Summers, R.M., 2019. Abnormal chest X-ray identification with generative adversarial one-class classifier. Proc. IEEE ISBI 1358–1361.
Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., Tyler, L.K., Henson, R.N., et al., 2017. The cambridge centre for ageing and neuroscience (cam-can) data repository: structural and functional mri, meg, and cognitive data from a cross-sectional adult lifespan sample. Neuroimage 144, 262–269.
Tustison, N.J., et al., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29, 1310–1320.
Vargas-Muñoz, J.E., Chowdhury, A.S., Alexandre, E.B., Galvão, F.L., Miranda, P.A.V., Falcão, A.X., 2019. An iterative spanning forest framework for superpixel segmentation. IEEE Trans. Image Process 28, 3477–3489.
Vasilakos, A.V., Tang, Y., Yao, Y., et al., 2016. Neural networks for computer-aided diagnosis in medicine: a review. Neurocomputing 216, 700–708.
Wang, L., et al., 2001. Statistical analysis of hippocampal asymmetry in schizophrenia. Neuroimage 14, 531–545.