

# HeMIS: Hetero-Modal Image Segmentation

Mohammad Havaei<sup>1</sup>(✉), Nicolas Guizard<sup>1</sup>, Nicolas Chapados<sup>1</sup>,  
and Yoshua Bengio<sup>2</sup>

<sup>1</sup> Imagia Inc., Montreal, QC, Canada

{mohammad.havaei,nicolas.guizard,nicolas.chapados}@imagia.com

<sup>2</sup> Université de Montréal, Montréal, Canada

yoshua.bengio@umontreal.ca

<http://www.imagia.com>

**Abstract.** We introduce a deep learning image segmentation framework that is extremely robust to missing imaging modalities. Instead of attempting to impute or synthesize missing data, the proposed approach learns, for each modality, an embedding of the input image into a single latent vector space for which arithmetic operations (such as taking the mean) are well defined. Points in that space, which are averaged over modalities available at inference time, can then be further processed to yield the desired segmentation. As such, any combinatorial subset of available modalities can be provided as input, without having to learn a combinatorial number of imputation models. Evaluated on two neurological MRI datasets (brain tumors and MS lesions), the approach yields state-of-the-art segmentation results when provided with all modalities; moreover, its performance degrades remarkably gracefully when modalities are removed, significantly more so than alternative mean-filling or other synthesis approaches.

**Keywords:** Segmentation · Multi-modal · Deep learning · Convolutional neural networks · Data abstraction · Data imputation

## 1 Introduction

In medical image analysis, image segmentation is an important task and is primordial to visualize and quantify the severity of the pathology in clinical practice. Multi-modality imaging provides complementary information to discriminate specific tissues, anatomies and pathologies. Numerous automatic approaches have been developed to speed up medical image segmentation such as *Multi-atlas based approaches* [4] and *model-based approaches* [12].

Both strategies are typically optimized for a specific set of multi-modal images and usually require these modalities to be available. In clinical settings, image acquisition and patient artifacts, among other hurdles, make it difficult to fully exploit all the modalities; as such, it is common to have one or more modalities to be missing for a given instance. This problem is not new, and the subject of missing data analysis has spawned an immense literature in statistics

(e.g. [13]). In medical imaging, a number of approaches have been proposed, some of which require to re-train a specific model with the missing modalities or to synthesize them [6]. Synthesis can improve multi-modal classification by adding information of the missing modalities in the context of a simple classifier such as random forests [11]. Approaches to imitate with fewer features a classifier trained with a complete set of features have also been proposed [7]. Nevertheless, it should stand to reason that a more complex model should be capable of extracting relevant features from just the available modalities without relying on artificial intermediate steps such as imputation or synthesis.

This paper proposes a deep learning framework (HeMIS) that can segment medical images from incomplete multi-modal datasets. Deep learning [3] has shown an increasing popularity in medical image processing for segmenting but also to synthesize missing modalities [11]. Here, the proposed approach learns, separately for each modality, an embedding of the input image into a latent space. In this space, arithmetic operations (such as computing first and second moments of a collection of vectors) are well defined and can be taken over the different modalities available at inference time. These computed moments can then be further processed to estimate the final segmentation. This approach presents the advantage of being robust to any combinatorial subset of available modalities provided as input, without the need to learn a combinatorial number of imputation models.

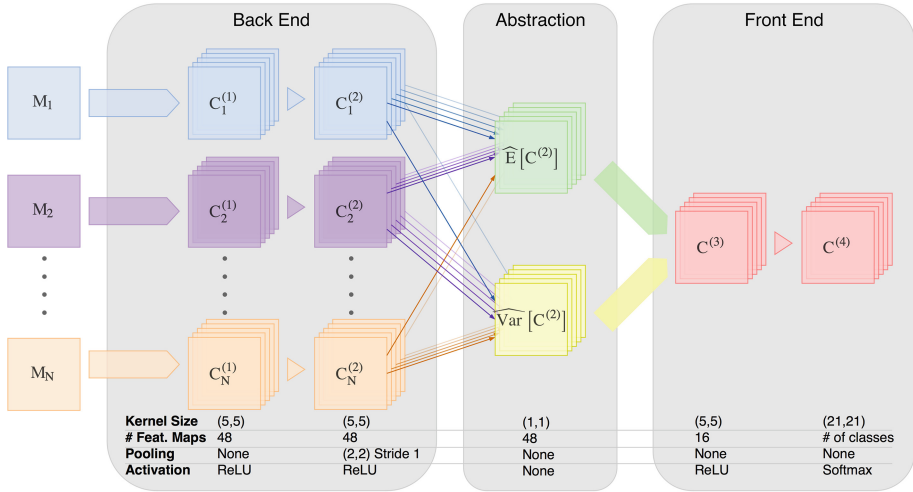
## 2 Method

### 2.1 Hetero-Modal Image Segmentation

Typical convolutional neural network (CNN) architectures take a multiplane image as input and process it through a sequence of convolutional layers (followed by nonlinearities such as  $\text{ReLU}(\cdot) \equiv \max(0, \cdot)$ ), alternating with optional pooling layers, to yield a per-pixel or per-image output [3]. In such networks every input plane is assumed to be present within a given instance: since the very first convolutional layer mixes input values coming from all planes, any missing plane introduces a bias in the computation that the network is not equipped to deal with.

We propose an approach wherein each modality is initially processed by its own convolutional pipeline, independently of all others. After a few independent stages, feature maps from all available modalities are *merged* by computing map-wise statistics such as the mean and the variance, quantities whose expectation does not depend on the number of terms (i.e. modalities) that are provided. After merging, the mean and variance feature maps are concatenated and fed into a final set of convolutional stages to obtain network output. This is illustrated in Fig. 1. In this procedure, each modality contributes a separate term to the mean and variance; in contrast to a vanilla CNN architecture, a missing modality does not throw this computation off: the mean and variance terms simply become estimated with larger uncertainty. In seeking to be robust to any subset of missing

modalities, we call this approach *hetero-modal* rather than multi-modal, recognizing that in addition to taking advantage of several modalities, it can take advantage of a diverse, instance-varying, set of modalities. In particular, it does not require that a “least common denominator” modality be present for every instance, as sometimes needed by common imputation methods.



**Fig. 1.** Illustration of the Hetero-Modal Image Segmentation architecture. Modalities available at inference time,  $M_k$ , are provided to independent modality-specific convolutional layers in the **back end**. Feature maps statistics (first & second moments) are computed in the **abstraction layer**, which after concatenation are processed by further convolutional layers in the **front end**, yielding pixelwise classifications outputs.

Let  $k \in \mathcal{K} \subseteq \{1, \dots, N\}$  denote a modality within the set of available modalities for a given instance, and  $M_k$  represent the image of the  $k$ -th modality. For simplicity, in this work we assume 2D data (e.g. a single slice of a tomographic image), but it can be extended in an obvious way to full 3D sections. As shown on Fig. 1, HeMIS proceeds in three stages:

- 1. Back End:** In our implementation, this consists of two convolutional layers with ReLU, the second followed with a (2, 2) max-pooling layer, denoted respectively  $C_k^{(1)}$  and  $C_k^{(2)}$ . To ensure that the output layer consists of the same number of pixels as the input image, the convolutions are zero-padded and the stride for all operations (including max-pooling) is 1. In particular, pooling with a stride of 1 *does not downsample*, but simply “thickens” the feature maps; this is found to add some robustness to the results. The number of feature maps in each layer is given in Fig. 1. Let  $C_{k,\ell}^{(j)}$  be the the  $\ell$ -th feature map of  $C_k^{(j)}$ .

2. *Abstraction Layer*: Modality fusion is computed here, as first and second moments across available modalities in  $C^{(2)}$ , separately for each feature map  $\ell$ ,

$$\widehat{E}_\ell [C^{(2)}] = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} C_{k,\ell}^{(2)} \quad \text{and} \quad \widehat{\text{Var}}_\ell [C^{(2)}] = \frac{1}{|\mathcal{K}| - 1} \sum_{k \in \mathcal{K}} \left( C_{k,\ell}^{(2)} - \widehat{E}_\ell [C^{(2)}] \right)^2,$$

with  $\widehat{\text{Var}}_\ell [C^{(2)}]$  defined to be zero if  $|\mathcal{K}| = 1$  (a single available modality).

3. *Front End*: Finally the front end combines the merged modalities to produce the final model output. In our implementation, we concatenate all  $\widehat{E} [C^{(2)}]$  and  $\widehat{\text{Var}} [C^{(2)}]$  feature maps, pass them through a convolutional layer  $C^{(3)}$  with ReLU activation, to finish with a final layer  $C^{(4)}$  that has as many feature maps as there are target segmentation classes. The pixelwise posterior class probabilities are given by applying a softmax function across the  $C^{(4)}$  feature maps, and a full image segmentation is obtained by taking the pixelwise most likely posterior class. No further postprocessing on the resulting segment classes (such as smoothing) is done.

### 3 Data and Implementation Details

We studied the HeMIS framework on two neurological pathologies: Multiple Sclerosis (MS) with the MS Grand Challenge (MSGC) and a large Relapsing Remitting MS (RRMS) cohort, as well as glioma with the Brain Tumor Segmentation (BRATS) dataset [8].

**MS MSGC**: The MSGC dataset [10] provides 20 training MR cases with manual ground truth lesion segmentation and 23 testing cases from the Boston Children’s Hospital (CHB) and the University of North Carolina (UNC). We downloaded<sup>1</sup> the co-registered T1W, T2W, FLAIR images for all 43 cases as well as the ground truth lesion mask images for the 20 training cases. While lesions masks for the 23 testing cases are not available for download, an automated system is available to evaluate the output of a given segmentation algorithm.

**RRMS**: This dataset is obtained from a multi-site clinical study with 300 RRMS patients (mean age 37.5 yrs, SD 10.0 yrs). Each patient underwent an MRI that included FLAIR, T1W, T2W and T1 post-contrast (T1C) images.

**BRATS**. The *BRATS-2015* dataset contains 220 subjects with high grade and 54 subjects with low grade tumors. Each subject contains four MR modalities (FLAIR, T1W, T1C and T2) and comes with a voxel level segmentation ground truth of 5 labels: *healthy*, *necrosis*, *edema*, *non-enhancing tumor* and *enhancing tumor*. As done in [8], we transform each segmentation map into 3 binary maps which correspond to 3 tumor categories, namely; *Complete* (which contains all tumor classes), *Core* (which contains all tumor subclasses except “edema”) and *Enhancing* (which includes the “enhanced tumor” subclass). For each binary map, the Dice Similarity Coefficient (DSC) is calculated [8].

<sup>1</sup> <http://www.nitrc.org/projects/msseg/>.

*BRATS-2013* contains two test datasets; Challenge and Leaderboard. The Challenge dataset contains 10 subjects with high grade tumors while the Leaderboard dataset contains 15 subjects with high grade tumors and 10 subject with low grade tumors. There are no ground truth provided for these datasets and thus quantitative evaluation can be achieved via an online evaluation system [8]. In our experiments we used Challenge and Leaderboard datasets to compare the HeMIS segmentation performance to the state-of-the-art, when trained on all modalities. To deal with class imbalance, we adopt the patch-wise training procedure mentioned in [5]. We make the HeMIS architecture robust to missing modalities by randomly dropping any number for a given training example. We refer to this training scheme as *pseudo-curriculum training*.

## 4 Experiments and Results

We first validate HeMIS performance against state-of-the-art segmentation methods on the two challenge datasets: MSGC and BRATS. Since the test data and the ranking table for BRATS 2015 are not available, we submitted results to BRATS 2013 challenge and leaderboard. These results are presented in Table 1.<sup>2</sup> As we observe, HeMIS outperforms Tustison et al. [12], the winner of the BRATS 2013 challenge, on most tumor region categories.

The MSGC dataset illustrates a direct application of HeMIS flexibility as only three modalities (T1W, T2W and FLAIR) are provided for a small training set. Therefore, given the small number of subjects, we first trained HeMIS on RRMS dataset with four modalities and fine-tuned on MSGC. Our results were submitted to the MSGC website<sup>3</sup>, with a results summary appearing in Table 2. The MSGC segmentation results include three other supervised approaches; when compared to them, HeMIS obtains highly competitive results with a combined score of 83.2%, where 90.0% would represent human performance given inter-rater variability.

The main advantage of HeMIS lies in its ability to deal with missing modalities, specifically when different subjects are missing different modalities. To illustrate the model’s flexibility in such circumstances, we compare HeMIS performance to two common approaches to deal with random missing modalities. The first, *mean-filling*, is to replace a missing modality by the modality’s mean value. In our case since all means are zero by construction, replacing a missing modality by zeros can be viewed as imputing with the mean. The second approach is to train a *multi-layer perceptron* (MLP) to predict the expected value of specific missing modality given the available ones. Since neural networks are generally trained for a unique task, we need to train 28 different MLPs (one for each  $\circ$  in Table 3 for a given dataset) to account for different possibilities of missing modalities. We used the same MLP architecture for all these models,

<sup>2</sup> Note that the results mentioned in Table 1 are from methods competing in the BRATS 2013 challenge for which a static table is provided at <https://www.virtualskeleton.ch/BraTS/StaticResults2013..>

<sup>3</sup> <http://www.ia.unc.edu/MSseg.>

**Table 1.** Comparison of HeMIS when trained on all modalities against BRATS-2013 Leaderboard and Challenge winners, in terms of Dice Similarity (scores from [8]).

Method	Leaderboard			Challenge		
	Complete	Core	Enhancing	Complete	Core	Enhancing
Tustison [12]	79	65	53	87	<b>78</b>	<b>74</b>
Zhao [14]	79	59	47	84	70	65
Meier [8]	72	60	53	82	73	69
HeMIS	<b>83</b>	<b>67</b>	<b>57</b>	<b>88</b>	75	<b>74</b>

**Table 2.** Results of the full dataset training on the MSGC. For each rater (CHB and UNC), we provide the volume difference (VD), surface distance (SD), true positive rate (TPR), false positive rate (FPR) and the method’s score as in [10].

Method	Rater	VD (%)	SD (mm)	TPR (%)	FPR (%)	Score
Souplet et al. [9]	CHB	86.4	8.4	58.2	70.6	80.0
	UNC	57.9	7.5	49.1	76.3	
Geremia et al. [2]	CHB	<b>52.4</b>	<b>5.4</b>	59.0	71.5	82.1
	UNC	<b>45.0</b>	<b>5.7</b>	51.2	76.7	
Brosch et al. [1]	CHB	63.5	7.4	47.1	<b>52.7</b>	<b>84.0</b>
	UNC	52.0	6.4	<b>56.0</b>	<b>49.8</b>	
HeMIS	CHB	127.4	7.5	<b>66.1</b>	55.3	83.2
	UNC	68.2	6.6	52.3	61.3	

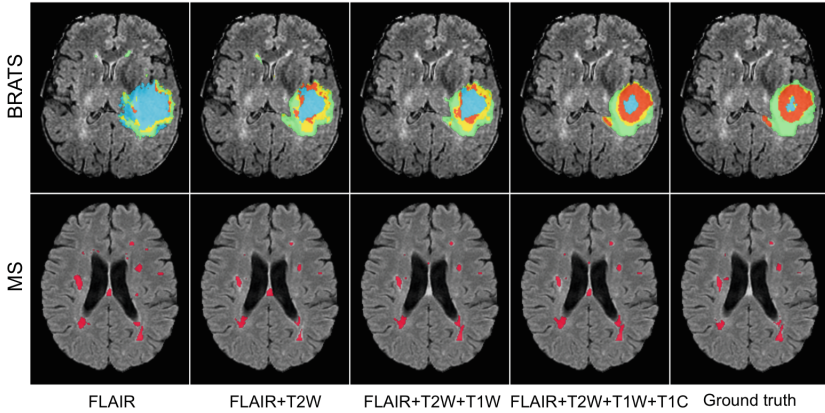
which consists of 2 hidden layers with 100 hidden units each, trained to minimize the mean squared error.

Table 3 shows the DSC for this experiment on the test set. On the BRATS dataset, for the Core category, HeMIS achieves the best segmentation in almost all cases (14 out of 15) and for the Complete and Enhancing categories it leads in most cases (10 and 9 cases out of 15 respectively). Also, the mean-filling approach hardly outperforms HeMIS or MLP-imputation. These results are consistent with the MS lesion segmentation dataset, where HeMIS outperforms other imputation approaches in 9 out of 15 cases. In scenarios where only one or two modalities are missing, while both HeMIS and MLP-imputation obtain good results, HeMIS outperforms the latter in most cases on both datasets. On BRATS, when missing 3 out of 4 modalities, HeMIS outperforms the MLP in a majority of cases. Moreover, whereas the HeMIS performance only gradually drops as additional modalities become missing, the performance drop for MLP-imputation and mean-filling is much more severe. On the RRMS cohort, the MLP-imputation appears to obtain slightly better segmentations when only one modality is available.

Although it is expected that tumor sub-label segmentations should be less accurate with fewer modalities, we should still hope for the model to report a

**Table 3.** Dice similarity coefficient (DSC) results on the RRMS and BRATS test sets (%). The table shows the DSC for different MRI modalities being either absent ( $\circ$ ) or present ( $\bullet$ ), in order of FLAIR ( $F$ ), T1W ( $T_1$ ), T1C ( $T_{1c}$ ), T2W ( $T_2$ ). Results are reported for HeMIS, Mean (mean-filling) and the imputation MLP.

Modalities					RRMS			BRATS							
					Lesion			Complete			Core			Enhancing	
$F$	$T_1$	$T_{1c}$	$T_2$	HeMIS	Mean	MLP	HeMIS	Mean	MLP	HeMIS	Mean	MLP	HeMIS	Mean	MLP
$\circ$	$\circ$	$\bullet$		1.74	2.66	<b>12.77</b>	58.48	2.70	<b>61.50</b>	<b>40.18</b>	4.00	37.32	<b>20.31</b>	6.25	18.62
$\circ$	$\circ$	$\bullet$	$\circ$	2.67	0.00	<b>3.51</b>	<b>33.46</b>	23.11	2.04	<b>44.55</b>	23.90	17.70	<b>49.93</b>	30.02	32.92
$\circ$	$\bullet$	$\circ$	$\circ$	3.89	0.00	<b>6.64</b>	<b>33.22</b>	0.00	2.07	<b>17.42</b>	0.00	10.52	4.67	6.25	<b>10.78</b>
$\bullet$	$\circ$	$\circ$	$\circ$	34.48	9.77	<b>38.46</b>	71.26	<b>72.30</b>	63.81	<b>37.45</b>	0.00	34.26	5.57	6.25	<b>15.90</b>
$\circ$	$\circ$	$\bullet$	$\bullet$	<b>27.52</b>	4.31	25.83	<b>67.59</b>	35.01	64.97	<b>63.39</b>	30.92	49.38	<b>65.38</b>	39.00	60.30
$\circ$	$\bullet$	$\bullet$	$\bullet$	8.21	0.00	<b>8.26</b>	<b>45.93</b>	23.63	1.99	<b>55.06</b>	41.89	26.55	<b>62.40</b>	43.80	40.93
$\bullet$	$\bullet$	$\bullet$	$\bullet$	38.81	11.62	<b>39.15</b>	<b>80.28</b>	75.58	78.13	<b>49.52</b>	0.00	48.97	22.26	6.25	<b>25.18</b>
$\circ$	$\bullet$	$\circ$	$\circ$	<b>31.25</b>	8.31	29.39	<b>69.56</b>	1.77	66.88	<b>47.26</b>	2.63	43.66	23.56	6.25	<b>26.37</b>
$\bullet$	$\circ$	$\circ$	$\bullet$	<b>39.64</b>	33.31	38.55	<b>82.1</b>	81.01	81.35	<b>53.42</b>	25.94	52.41	23.19	6.25	<b>25.01</b>
$\bullet$	$\circ$	$\bullet$	$\circ$	<b>41.38</b>	6.42	39.33	79.8	45.97	<b>81.13</b>	<b>66.12</b>	29.85	65.51	<b>67.12</b>	35.14	66.19
$\bullet$	$\bullet$	$\bullet$	$\circ$	<b>41.97</b>	9.00	40.63	80.88	81.57	<b>82.19</b>	69.26	62.13	<b>69.34</b>	<b>71.30</b>	67.13	70.93
$\bullet$	$\bullet$	$\circ$	$\bullet$	<b>46.6</b>	41.12	41.83	<b>83.87</b>	77.84	80.40	<b>57.76</b>	20.66	53.46	<b>28.46</b>	6.25	28.34
$\bullet$	$\circ$	$\bullet$	$\bullet$	<b>41.90</b>	38.95	41.47	82.78	64.19	<b>83.37</b>	<b>70.62</b>	42.36	70.45	70.52	49.62	<b>70.56</b>
$\circ$	$\bullet$	$\bullet$	$\bullet$	<b>34.98</b>	5.78	29.46	<b>70.98</b>	30.86	67.85	<b>66.60</b>	45.79	55.40	<b>67.84</b>	50.21	64.81
$\bullet$	$\bullet$	$\bullet$	$\bullet$	<b>48.66</b>	43.48	43.48	<b>83.15</b>	82.43	82.43	<b>72.5</b>	71.46	71.46	<b>75.37</b>	72.08	72.08
# Wins / 15				<b>9</b>	0	6	<b>10</b>	1	4	<b>14</b>	0	1	<b>9</b>	0	6



**Fig. 2.** HeMIS segmentation results on BRATS and MS subjects. For BRATS (first row) the segmentation colors describe necrosis (blue), non-enhancing (yellow), active core (orange) and edema (green). For the MS case, the lesions are highlighted in red. The columns present the results overlaid on top of a FLAIR image for different combinations of input modalities, with ground truth in the last column.

sensible characterization of the tumor “footprint”. While MLP and mean-filling fail in this respect, HeMIS quite well achieves this goal by outperforming in almost all cases of the Complete and Core tumor categories. This can also be

seen in Fig. 2 where we show how adding modalities to HeMIS improves its ability to achieve a more accurate segmentation. From Table 3, we can also infer that the FLAIR modality is the most relevant for identifying the Complete tumor while TIC is the most relevant for identifying Core and Enhancing tumor categories. On the RRMS dataset, HeMIS results are also seen to degrade slower than the other imputation approaches, preserving good segmentation when modalities go missing. Indeed, as seen in Fig. 2, even though with FLAIR alone HeMIS already produces good segmentations, it is capable of further refining its results when adding modalities, by removing false positives and improving outlines of the correctly identified lesions or tumor.

## 5 Conclusion

We have proposed a new fully automatic segmentation framework for heterogeneous multi-modal MRI using a specialized convolutional deep neural network. The embedding of the multi-modal CNN back-end allows to train and segment datasets with missing modalities. We carried out an extensive validation on MS and glioma and achieved state-of-the-art segmentation results on two challenging neurological pathology image processing tasks. Importantly, we contrasted the graceful performance degradation of the proposed approach as modalities go missing, compared with other popular imputation approaches, which it achieves without requiring training specific models for every potential missing modality combination. Future work should concentrate on extending the approach to broader modalities outside of MRI, such as CT, PET and ultrasound.

## References

1. Brosch, T., Yoo, Y., Tang, L.Y.W., Li, D.K.B., Traboulsee, A., Tam, R.: Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 3–11. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-24574-4\\_1](https://doi.org/10.1007/978-3-319-24574-4_1)
2. Geremia, E., Menze, B.H., Ayache, N.: Spatially adaptive random forests, pp. 1344–1347 (2013)
3. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT Press, Cambridge (2016)
4. Guizard, N., Coupé, P., Fonov, V.S., Manjón, J.V., Arnold, D.L., Collins, D.L.: Rotation-invariant multi-contrast non-local means for ms lesion segmentation. *NeuroImage Clin.* **8**, 376–389 (2015)
5. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. arXiv preprint (2015). [arXiv:1505.03540](https://arxiv.org/abs/1505.03540)
6. Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., et al.: MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and atlas registration. *J. Nucl. Med.* **49**(11), 1875–1883 (2008)
7. Hor, S., Moradi, M.: Scandent Tree: a random forest learning method for incomplete multimodal datasets. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 694–701. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-24553-9\\_85](https://doi.org/10.1007/978-3-319-24553-9_85)



8. Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J.E.A.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE TMI* **34**(10), 1993–2024 (2015)
9. Souplet, J., Lebrun, C., Ayache, N., Malandain, G.: An automatic segmentation of T2-FLAIR multiple sclerosis lesions, 7 2008
10. Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., Warfield, S.: 3D segmentation in the clinic: A grand challenge ii: Ms lesion segmentation. *MIDAS* **2008**, 1–6 (2008)
11. Tulder, G., Bruijne, M.: Why does synthesized data improve multi-sequence classification? In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9349, pp. 531–538. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-24553-9\\_65](https://doi.org/10.1007/978-3-319-24553-9_65)
12. Tustison, N.J., Shrinidhi, K., Wintermark, M., Durst, C.R., Kandel, B.M., Gee, J.C., Grossman, M.C., Avants, B.B.: Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. *Neuroinformatics* **13**(2), 209–225 (2015)
13. Van Buuren, S.: *Flexible imputation of missing data*. CRC Press, Boca Raton (2012)
14. Zhao, L., Wu, W., Corso, J.J.: Semi-automatic brain tumor segmentation by constrained MRFs using structural trajectories. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013*. LNCS, vol. 8151, pp. 567–575. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40760-4\\_71](https://doi.org/10.1007/978-3-642-40760-4_71)