Multi-sequence learning for multiple sclerosis lesion segmentation in spinal cord MRI

Ricky Walsh, Malo Gaubert, Cédric Meurée, Burhan Rashid Hussein, Anne Kerbrat, Romain Casey, Benoît Combès, Francesca Galassi

Multi-sequence learning for multiple sclerosis lesion segmentation in spinal cord MRI

Ricky Walsh[1], Malo Gaubert[1,2], Cédric Meurée[1], Burhan Rashid Hussein[1], Anne Kerbrat[1,3], Romain Casey[4], Benoit Combès[1,†], and Francesca Galassi[1,†]

[1] Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Empenn, Rennes, France
[2] Department of Neuroradiology, Rennes University Hospital, Rennes, France
[3] Department of Neurology, Rennes University Hospital, Rennes, France
[4] Univ Lyon, Université Claude Bernard Lyon 1, Hospices Civils de Lyon, Fondation EDMUS, OFSEP, Centre de Recherche en Neurosciences de Lyon, Lyon, France
{ricky.walsh,benoit.combes,francesca.galassi}@inria.fr
[†] Equal contribution

Abstract. Automated tools developed to detect multiple sclerosis lesions in spinal cord MRI have thus far been based on processing single MR sequences in a deep learning model. This study is the first to explore a multi-sequence approach to this task and we propose a method to address inherent issues in multi-sequence spinal cord data, i.e., differing fields of view, inter-sequence alignment and incomplete sequence data for training and inference. In particular, we investigate a simple missing-modality method of replacing missing features with the mean over the available sequences. This approach leads to better segmentation results when processing a single sequence at inference than a model trained directly on that sequence, and our experiments provide valuable insights into the mechanism underlying this surprising result. In particular, we demonstrate that both the encoder and decoder benefit from the variability introduced in the multi-sequence setting. Additionally, we propose a latent feature augmentation scheme to reproduce this variability in a single-sequence setting, resulting in similar improvements over the single-sequence baseline.

Keywords: Multiple Sclerosis · Missing Modality · Segmentation.

1 Introduction

Automated tools using deep learning have been developed in recent years to aid in detecting and delineating Multiple Sclerosis (MS) lesions in Spinal Cord (SC) MRI [4, 10, 6]. However, these tools process single input images per subject, rather than combining multiple MR sequences, as are often available [13]. While multi-modal brain MRI segmentation is common [1], spinal MRI poses significantly greater challenges due to variations in fields of view, misalignments between acquisitions, and potential motion artifacts [8].

Accepted to the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2024

Moreover, the diversity of protocols for SC acquisitions results in a "missing modality" problem, i.e., not every sequence is available for each subject. This is an important issue in deep learning for medical image and has been studied over the last decade. A seminal 2016 study [5] proposed processing each modality separately, and taking the mean and variance of feature maps across the available modalities. However, this approach risks losing modality-specific information. Recent works focus on learning better features for each modality, using a combination of modality-specific encoders and shared encoders [11], or using distillation to guide modality-specific encoders with a more informative "teacher modality" [12]. Both [11, 12] use a conceptually simple approach to fusing the features from each modality, where the feature maps of missing modalities are substituted by the mean of the features across available modalities (*Mean Imputation*). Both studies outperform several state-of-the-art missing-modality techniques in brain tumour segmentation.

The contributions of this study are as follows. Firstly, it is the first work, to our knowledge, to explore a multi-sequence approach for automated MS lesion segmentation in SC MRI. As such, we propose a pre-processing pipeline to deal with the inherent challenges, including varying fields of view and inter-sequence misalignments. Secondly, we carry out a thorough analysis into the unexpected positive impact of *Mean Imputation* on single-sequence inference, revealing that the benefit arises from an improved robustness of both the encoder and decoder. Finally, we propose a latent feature augmentation scheme for single-sequence training to replicate the benefit of variability observed in multi-sequence training.

2 Method

2.1 Data

The dataset used in this study consisted of MR images of 247 subjects with MS from clinical studies[1] and the French MS Registry (OFSEP[2])[9]. All subjects provided written consent. The study was approved by the relevant ethics committee and is compliant with French data confidentiality regulations.

Depending on the subject, different combinations of sequences among sagittal T2-w (T2Sag), sagittal STIR, axial T2*-w (T2*Ax), axial T2-w (T2Ax), sagittal T1-w (T1Sag) without gadolinium, and MP2RAGE were acquired to image the SC. For each subject, we included all the available acquisitions.

The fields of view (FOV) vary significantly between different sequences for a given subject. For example, the T2Sag scans (acquired in two overlapping slabs) can cover from the brainstem to the lumbar vertebrae, while T2*Ax only covers C1 to C7. All subjects had at least one T2Sag scan of the upper SC (typically to T3-T4), which is the most common acquisition in clinical practice [13]. For 102 subjects, the acquisitions did not include the lower SC.

[1] clinicaltrials.gov study IDs: NCT02117375, NCT04220814, NCT04918225
[2] OFSEP data available on request: https://www.ofsep.org/en/data-access

For a given subject, SC lesions were delineated on the T2Sag image by one of six experts using all available acquisitions. An experienced rater revised all the obtained segmentation masks to increase consistency across the dataset. The annotated dataset was split into a training set (N=162 subjects), a validation set (N=27) and test set (N=58). The test set contained three subsets ($\mathcal{D}_A$, $\mathcal{D}_B$, $\mathcal{D}_C$) from distinct cohorts (see Supple. Fig 1 for more details).

2.2 Pre-processing

Our pre-processing pipeline consisted of the following four steps.

Intra-sequence fusion and alignment For a given subject, when multiple scans were acquired with one sequence (e.g. C1-C3 and C4-C7 for T2*Ax), these were merged into one, taking half of the overlapping region from each image. An initial rigid registration was performed between the upper and lower images of the same sequence to address artifacts resulting from misalignment. This registration was, however, not applied to the axial acquisitions, mostly because of the lack of overlap between images.

Inter-sequence alignment The SC is a narrow and highly mobile structure, so small subject movements between two acquisitions can introduce non-linear deformations. Correcting for subtle deformations between images of different sequences in a fully automated way is challenging. Moreover, non-linear registration is likely to significantly alter the intensity profile and thus can deteriorate the signal of interest. Therefore, we did not apply non-linear inter-sequence registration, but rather relied on aligning the SC centreline in each image. The centreline was found independently for all sequences using the corresponding *deepseg* models from the Spinal Cord Toolbox [2].

Inter-sequence resampling To address variations in image spacing and FOV, all available images were resampled to a reference image grid with isotropic 0.5mm spacing. The extent of this reference grid in the superior-inferior axis was determined by the maximum extent of the SC masks obtained in the previous step, plus a margin of 10mm either end to ensure the full SC was captured. Images were padded with zero for voxels outside of their initial coverage.

Data cropping Finally, to reduce image size and, thereby, the time required to train a model, the images were cropped and shifted around the SC centreline, as in [4]. Specifically, for each axial slice, the image was cropped to 48×48 voxels around the centreline. The cropped axial slices were then stacked vertically.

2.3 Architecture

T2Sag Baseline Our baseline 3D U-Net architecture, based on nnU-Net [7], processed a single sequence. Patch size was set to (48, 48, 320) voxels in RPI

orientation, covering the full pre-processed image in the axial plane. Four downsampling operations were employed and the convolution channels for each resolution were: 32, 64, 128, 256, 320. Deep supervision was applied to the two decoder layers closest to the original resolution. We used the sum of cross-entropy and Dice losses. The learning rate was 0.01 and was decayed polynomially. The SGD optimiser used Nesterov momentum (0.99) and weight decay (3×10^{-5}).

Multi-sequence When training with multiple sequences, one encoder was initialised per sequence, each generating a set of features with 320 channels, denoted as $\mathbf{f}_k$, where $k \in \mathcal{K}^* = \{\text{T2Sag, STIR, T2*Ax, T2Ax, T1Sag, MP2RAGE}\}$. These features were concatenated, resulting in a combined feature map of $320 \times n$ channels, where n is the number of sequences used. A convolution with 320 output channels was then applied, followed by a single decoder. Results of training models on specific combinations of sequences are presented in Sec. 3.1.

We merged features from different sequences only at the bottleneck of the U-Net model to mitigate the effect of potential remaining inter-sequence misalignments. The effective resolution at the bottleneck was $4 \times 4 \times 8$mm, whereas the input images had a finer $0.5 \times 0.5 \times 0.5$mm resolution. Similarly, the skip connections to the decoder were taken only from the T2Sag encoder.

Mean Imputation The missing-modality approach involved training a model on images from all subjects and all six sequences. The same architecture was used as in the multi-sequence setting, here with six encoders (see Supple. Fig. 2). The bottleneck had $320 \times 6 = 1,920$ channels after concatenating the features from each sequence. During both training and inference, if a sequence was missing for a subject, the 320 bottleneck features for that sequence were imputed by the mean of the features from the available sequences for that subject, as in [12]. Given a subject S_i, whose available sequences are denoted $\mathcal{K}_i \subset \mathcal{K}^*$, and missing sequences $\mathcal{J}_i = \mathcal{K}^* \setminus \mathcal{K}_i$, then the missing features $\mathbf{f}_{i,j}$, $j \in \mathcal{J}_i$, are imputed as

$$\mathbf{f}_{i,j} = \frac{1}{|\mathcal{K}_i|} \sum_{k \in \mathcal{K}_i} \mathbf{f}_{i,k} \tag{1}$$

2.4 Understanding Mean Imputation

Improved Encoder Results with *Mean Imputation* showed that training with multiple sequences benefited the model even when using a single sequence for inference, by learning more robust encoder and/or decoder representations. To isolate the effect on the decoder, we conducted an experiment with *Mean Imputation* with a **frozen encoder**. Specifically, we extracted the learned weights from the T2Sag Baseline model and froze these during training so that the *Mean Imputation* model could not learn better T2Sag features.

Secondly, we experimented with a **distillation loss** on bottleneck features as in [12]. We maintained T2Sag as the teacher modality, being the most common sequence in our dataset and the reference image for manual segmentations. The

distillation loss was calculated as L1 loss between the feature values of T2Sag and other sequences at the bottleneck. Given a subject S_i, whose available sequences are denoted $\mathcal{K}_i \subset$ {T2Sag, STIR, T2*Ax, T2Ax, T1Sag, MP2RAGE}, the distillation loss for this subject is given by

$$\mathcal{L}_{L1}(S_i) = \frac{1}{|\mathcal{K}_i| - 1} \sum_{k \in \mathcal{K}_i, k \neq T2Sag} ||\mathbf{f}_{i,T2Sag} - \mathbf{f}_{i,k})||_1 \quad (2)$$

The distillation loss was weighted by $\alpha = 0.1$, the optimal value found by [12], and added to the total loss.

Improved Decoder Results suggested the increased variability of features seen when training with *Mean Imputation* may benefit the generalisation of the decoder. To model this variability in a single-sequence setting, we propose a method applying random perturbations to the latent features of a T2Sag model. These perturbations were carefully determined based on the distribution of differences between T2Sag features and other sequences from a trained *Mean Imputation* model. We observed two parts to the distribution, which were modelled with a logistic distribution ($\mu = 0$, $s = 0.002$) for 25% of the differences and a normal distribution ($\mu = 0$, $\sigma = 0.4$) for the remaining 75%. Noise was generated according to this distribution during each training iteration and added to the bottleneck features from the T2Sag image ($\mathbf{f}_{T2Sag}$), followed by a Leaky ReLU layer to maintain original feature scale.

2.5 Evaluation Metrics

We used the Dice coefficient and lesion-wise F1 score to evaluate the models. The definition of lesion-wise F1 followed that of [10]: a ground-truth lesion is a true positive if at least 10% of its voxels are detected, while a predicted lesion is a false positive if over 70% of its voxels do not overlap with a ground-truth lesion. Lesion-wise sensitivity, precision and F1 can then be computed (F1 defined as $F1 = 2 \times (Precision \times Sensitivity)/(Precision + Sensitivity)$. Dice and F1 were computed for each image, and the mean result across images is reported. Further metrics are presented in Supple. Figs. 3 and 4.

3 Results and Discussion

3.1 Comparison to Baselines

Table 1 presents the lesion-wise F1 and Dice scores of baseline models trained on specific combinations of sequences, as well as the *Mean Imputation* technique. Since not every sequence was available for each subject, training on specific combinations of sequences means training on a subset of subjects. As a result, these models generally underperform compared to the T2Sag baseline, except for one instance where the model trained on T2Sag, T2*Ax and T1Sag achieves a slightly higher F1 (0.593) than the T2Sag baseline (0.583).

Table 1. The first four rows are models trained and tested on specific sequence combinations. *Mean Imputation* was trained on all sequences but evaluated with a subset at inference. Test sets $\mathcal{D}_A$, $\mathcal{D}_B$, and $\mathcal{D}_C$ are from distinct cohorts, and blank entries indicate missing sequence combinations. The best result for each test set is in bold.

	Sequences at Inference						Lesion-wise F1			Dice Coefficient		
Model	T2Sag	STIR	T2*	T2Ax	T1	MP2	$\mathcal{D}_A$	$\mathcal{D}_B$	$\mathcal{D}_C$	$\mathcal{D}_A$	$\mathcal{D}_B$	$\mathcal{D}_C$
Specific Sequences	•	○	○	○	○	○	0.593	0.517	0.583	0.452	0.403	0.469
	•	•	○	○	○	○	0.558	0.477	–	0.418	0.355	–
	•	○	•	○	•	○	–	–	0.593	–	–	0.445
	•	○	•	•	○	○	0.550	–	–	0.380	–	–
Mean Imputation	•	○	○	○	○	○	0.641	0.563	0.606	0.466	0.414	0.493
	•	•	○	○	○	○	0.637	**0.570**	–	0.470	**0.432**	–
	•	○	•	○	•	○	–	–	**0.622**	–	–	**0.518**
	•	○	•	•	○	○	**0.650**	–	–	**0.480**	–	–
	•	•	•	•	○	•	0.646	–	–	0.475	–	–

The *Mean Imputation* method was trained on all sequences and can adapt to the sequences available at inference. Incorporating additional sequences at inference generally improved performance compared to using T2Sag alone, although there was a slight decrease in F1 score when using T2Sag and STIR (0.637) vs. T2Sag alone (0.641) on test set $\mathcal{D}_A$. However, these differences in performance are modest. The largest improvements in F1 and Dice of 0.016 and 0.025, respectively, were observed when using T2Sag, T2*Ax and T1Sag on test set $\mathcal{D}_C$ compared to using T2Sag alone. Future reasearch aims to improve performance in the multi-sequence setting by, for example, fusing at different depths, investigating alternative fusion methods, and leveraging manual segmentations from other sequences as auxiliary tasks for feature extraction.

Comparing *Mean Imputation* to the T2Sag baseline yields a more interesting finding. Surprisingly, *Mean Imputation* significantly outperforms the T2Sag baseline, even when using only T2Sag data at inference. Both models were trained on the same amount of T2Sag data, with the same hyperparameters and have nearly identical architectures, differing only in the bottleneck. *Mean Imputation* replaces the features of missing sequences with the mean of the features across the available sequences; when only T2Sag is available, the T2Sag features are simply replicated six times. The increased number of parameters at the bottleneck in the *Mean Imputation* model cannot alone account for the improvement over the T2Sag baseline. Training a model on a single sequence, replicating the 320 feature channels and passing the concatenated features through a convolution is mathematically equivalent to the T2Sag Baseline method, given the linearity of the convolution operation. This suggests that additional factors are contributing to the improved performance of *Mean Imputation*.

3.2 Understanding Mean Imputation

We propose two explanations for the improvement of *Mean Imputation* over the T2Sag baseline. First, the approach may improve the quality of the features generated by the T2Sag encoder. Second, the variability introduced in the latent features may aid the fusion layer and decoder to learn more robust representations. The following experiments explore these two potential mechanisms.

Improved Encoder If a sequence is missing during training then the weights in the bottleneck fusion layer associated to that sequence will be applied to the mean features from all available sequences, including T2Sag. In this way, the feature representations generated by the different sequence-specific encoders are implicitly guided to be similar; if these representations were dissimilar, the sequence-specific fusion weights could not compensate when the sequence is missing. Such regularisation may improve the T2Sag feature learning process.

To validate this hypothesis, we conducted an experiment by training the *Mean Imputation* method while freezing the T2Sag encoder, as described in Sec. 2.4. Table 2 shows that this indeed leads to a drop in lesion-wise F1 from 0.603 to 0.591, suggesting that better T2Sag features play a role in the improved performance. However, the score remains above that of the T2Sag baseline (F1=0.564), implying that there are other contributory factors. Future research will explore replicating this effect in a single-sequence setting, e.g., by encouraging consistent features for the same image under various input augmentations, as in some self-supervised methods [3].

The authors of [12] propose a distillation loss to learn better features for less informative modalities by aligning them with a teacher modality. We hypothesise that this loss may also regularise the encoder of the teacher modality and, as such, may therefore achieve a similar effect to the mechanism of learning better T2Sag features discussed above. However, when applying the distillation loss, we observe a decrease in lesion-wise F1 from 0.603 to 0.575 and a drop in Dice from 0.457 to 0.435.

The lower performance with distillation loss contrasts to the benefit observed in [12]. This difference may be due to several factors. Firstly, the optimal value of $\alpha = 0.1$ in [12] might not be the ideal value for our setting. However, all tested values of $\alpha \in (0, 1]$ in [12] showed significant improvements over $\alpha = 0$, i.e., no distillation. Secondly, that study demonstrated a benefit only when using the least informative modality for inference, so it is unclear how the loss affected the strongest modality. Finally, using an L1 loss may not be suitable due to variations in FOV; for example, T2*Ax does not cover the thoracic region, so applying the L1 loss to features in this region may yield unexpected results.

Improved Decoder The *Frozen Encoder* results indicate that some of the benefit of *Mean Imputation* stems from an improved bottleneck and decoder. This is likely due to greater diversity of features seen by the fusion layer, enabling more robust weights to be learned. We explore if this benefit can be recreated without additional sequences, by introducing perturbations to the latent features, as

Table 2. Results using only T2Sag at inference (mean ± std. dev.). Each model was trained three times with different initialisation. A Wilcoxon signed-rank test compared the results for each image to T2Sag Baseline – Significance Levels: * 0.1; ** 0.05.

Method	Lesion-wise F1	Dice Coefficient
T2Sag Baseline	0.564 ± 0.031	0.441 ± 0.011
T2Sag – Feature Augmentation	0.586 ± 0.026 **	0.446 ± 0.014 *
Mean Imputation – Frozen Encoder	0.591 ± 0.022 **	0.448 ± 0.015 *
Mean Imputation – Distillation Loss	0.575 ± 0.017 *	0.435 ± 0.010
Mean Imputation	**0.603** ± 0.025 **	**0.457** ± 0.013 **

described in Sec. 2.4. Table 2 shows that applying these perturbations improves lesion-wise F1 over the T2Sag baseline (0.586 vs. 0.564), similar to training *Mean Imputation* with a frozen encoder (0.591). This suggests a similar improvement to the robustness of the decoder can be achieved with this latent augmentation.

3.3 Comparison to Other Studies

[4] reported a median Dice score of 0.576 for their T2-w model, whereas the median Dice of *Mean Imputation* using T2Sag for inference was 0.497. The median lesion-wise F1 for *Mean Imputation* was 0.667 with lesion-wise sensitivity and precision of 77% and 69%, respectively, which is a similar precision but lower than the sensitivity of 90% reported in [4]. However, results for the same method can vary significantly on different data, as is evident in Table 1. Furthermore, the T2-w data of [4] includes axial and isotropic data along with T2Sag, which may contribute to better segmentation. Finally, the F1 and Dice scores of *Mean Imputation* were similar to those observed in [10] comparing radiologists to an adjudicated ground truth (F1=0.667, Dice=0.489), and the higher sensitivity of our method (77% vs. 50%) indicates that this model could be a useful aid.

4 Conclusion

This study is the first to explore a multi-sequence approach to SC MS lesion segmentation in MRI. We proposed a pre-processing and modelling pipeline to address the challenges inherent in multi-sequence SC MRI data. Using multiple sequences for inference yielded some improvements, albeit modest, over using T2Sag alone. However, training with multiple sequences in a missing-modality setting led to a significant improvement, even when using only one sequence for inference. Our experiments demonstrated that replacing missing features during training with the mean over available sequences helped to regularise both the encoder and decoder. Finally, based on our findings, we proposed a method of applying augmentation to the latent features while training a single-sequence model, replicating the benefit of variability observed in the multi-sequence setting. Our pipeline, analysis, and experimental results collectively advance the field of automated MS lesion segmentation in spinal cord MRI.

Acknowledgements This study was supported by the French National Research Agency (Agence Nationale de la Recherche, ANR) within the France 2030 program (ANR-21-RHUS-0014) and the French doctoral program in artificial intelligence (ANR-20-THIA-0018). Data collection was supported by a grant from the French Ministry of Health and handled by ANR within the France 2030 program (ANR-10-COHO-002 OFSEP).

References

1. Danelakis, A., Theoharis, T., Verganelakis, D.A.: Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. Computerized Medical Imaging and Graphics **70** (Dec 2018). https://doi.org/10.1016/j.compmedimag.2018.10.002
2. De Leener, B., Lévy, S., Dupont, S.M., Fonov, V.S., Stikov, N., Louis Collins, D., Callot, V., Cohen-Adad, J.: SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. NeuroImage **145**(Pt A) (Jan 2017). https://doi.org/10.1016/j.neuroimage.2016.10.009
3. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., Piot, B., kavukcuoglu, k., Munos, R., Valko, M.: Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning. In: Advances in Neural Information Processing Systems. vol. 33 (2020)
4. Gros, C., De Leener, B., Badji, A., Maranzano, J., Eden, D., Dupont, S.M., Talbott, J., Zhuoquiong, R., Liu, Y., Granberg, T., Ouellette, R., Tachibana, Y., Hori, M., Kamiya, K., Chougar, L., Stawiarz, L., Hillert, J., Bannier, E., Kerbrat, A., Edan, G., Labauge, P., Callot, V., Pelletier, J., Audoin, B., Rasoanandrianina, H., Brisset, J.C., Valsasina, P., Rocca, M.A., Filippi, M., Bakshi, R., Tauhid, S., Prados, F., Yiannakas, M., Kearney, H., Ciccarelli, O., Smith, S., Treaba, C.A., Mainero, C., Lefeuvre, J., Reich, D.S., Nair, G., Auclair, V., McLaren, D.G., Martin, A.R., Fehlings, M.G., Vahdat, S., Khatibi, A., Doyon, J., Shepherd, T., Charlson, E., Narayanan, S., Cohen-Adad, J.: Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. NeuroImage **184** (Jan 2019). https://doi.org/10.1016/j.neuroimage.2018.09.081
5. Havaei, M., Guizard, N., Chapados, N., Bengio, Y.: HeMIS: Hetero-Modal Image Segmentation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_54
6. Hussein, B.R., Meurée, C., Gaubert, M., Masson, A., Kerbrat, A., Combès, B., Galassi, F.: A Study on Loss Functions and Decision Thresholds for the Segmentation of Multiple Sclerosis Lesions on Spinal Cord MRI. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI) (Apr 2023). https://doi.org/10.1109/ISBI53787.2023.10230676
7. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2) (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z
8. Moccia, M., Ruggieri, S., Ianniello, A., Toosy, A., Pozzilli, C., Ciccarelli, O.: Advances in spinal cord imaging in multiple sclerosis. Therapeutic Advances in Neurological Disorders **12** (Jan 2019). https://doi.org/10.1177/1756286419840593

9. Vukusic, S., Casey, R., Rollot, F., Brochet, B., Pelletier, J., Laplaud, D.A., De Seze, J., Cotton, F., Moreau, T., Stankoff, B.: Observatoire Français de la Sclérose en Plaques (OFSEP): A unique multimodal nationwide MS registry in France. Multiple Sclerosis Journal **26**(1) (2020)
10. Walsh, R., Meuree, C., Kerbrat, A., Masson, A., Hussein, B.R., Gaubert, M., Galassi, F., Combes, B.: Expert Variability and Deep Learning Performance in Spinal Cord Lesion Segmentation for Multiple Sclerosis Patients. In: 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS) (Jun 2023). https://doi.org/10.1109/CBMS58004.2023.00263
11. Wang, H., Chen, Y., Ma, C., Avery, J., Hull, L., Carneiro, G.: Multi-Modal Learning With Missing Modality via Shared-Specific Feature Modelling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
12. Wang, H., Ma, C., Zhang, J., Zhang, Y., Avery, J., Hull, L., Carneiro, G.: Learnable Cross-modal Knowledge Distillation for Multi-modal Learning with Missing Modality. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Lecture Notes in Computer Science, Cham (2023). https://doi.org/10.1007/978-3-031-43901-8_21
13. Wattjes, M.P., Ciccarelli, O., Reich, D.S., Banwell, B., de Stefano, N., Enzinger, C., Fazekas, F., Filippi, M., Frederiksen, J., Gasperini, C., Hacohen, Y., Kappos, L., Li, D.K.B., Mankad, K., Montalban, X., Newsome, S.D., Oh, J., Palace, J., Rocca, M.A., Sastre-Garriga, J., Tintoré, M., Traboulsee, A., Vrenken, H., Yousry, T., Barkhof, F., Rovira, À., Wattjes, M.P., Ciccarelli, O., de Stefano, N., Enzinger, C., Fazekas, F., Filippi, M., Frederiksen, J., Gasperini, C., Hacohen, Y., Kappos, L., Mankad, K., Montalban, X., Palace, J., Rocca, M.A., Sastre-Garriga, J., Tintore, M., Vrenken, H., Yousry, T., Barkhof, F., Rovira, A., Li, D.K.B., Traboulsee, A., Newsome, S.D., Banwell, B., Oh, J., Reich, D.S., Oh, J.: 2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. The Lancet Neurology **20**(8) (Aug 2021). https://doi.org/10.1016/S1474-4422(21)00095-8