Prediction of Progression in Multiple Sclerosis Patients

Adrian Tousignant
McGill University
adrian.tousignantduran@mail.mcgill.ca

Doina Precup
McGill University
dprecup@cs.mcgill.ca

Tal Arbel
McGill University
arbel@cim.mcgill.ca

Abstract

This paper presents the first work to explore automatic end-to-end deep learning prediction of future patient disability progression (one year from baseline) based on multi-modal brain Magnetic Resonance Images (MRI) of patients with Multiple Sclerosis (MS), a neurological disease characterized by inflammation and degeneration of the central nervous system. The model is trained on two large proprietary, multi-scanner, multi-center, clinical trial datasets of patients with Relapsing-Remitting Multiple Sclerosis (RRMS). Results illustrate that the network can accurately predict future disease progression, measured by the change in the extended disability status scale (EDSS) score over time, based on multi-modal MRI and associated lesion labels. Using data from 465 patients, we achieve an AUC of 0.66 ± 0.055 when training on patients' MRI modalities and 0.701 ± 0.027 when supplementing the model with lesion label data.

1 Introduction

The recent successes of deep convolutional neural networks (CNN) in image classification challenges such as ImageNet [1] have sparked a lot of interest from the computer vision community. Although deep learning techniques have shown promising results in lesion segmentation in Multiple Sclerosis (MS) [2, 3, 4], automatic prediction of disease worsening, or progression, in MS remains a complex open problem. Challenges presented to deep learning methods applied to medical image analysis include the loss of fine-grain features as the information propagates through the encoding and pooling layers. In this work, we present the first end-to-end 3D CNN trained on patients' baseline MRI predicting progression for MS patients within the year following baseline. Our method proposes to avoid fine grain feature loss through the use of parallel convolutional pathways with varying resolution levels. In addition, we investigate the effect of training the network with supplementary information provided by T2-weighted and gadolinium-enhanced lesions, well-known biomarkers of disease activity [5, 6, 7]. Finally, we explore how Monte Carlo (MC) dropout [8] can be used as an uncertainty estimate to improve our choice of operating point.

2 Methodology

Our model consists of three consecutive 3D convolutional blocks followed by 2 fully connected layers. Each convolutional block is composed of parallel convolutional layers of varying kernel size to extract features at different levels of resolution. The last layer is then flattened and fed to two successive fully connected layers with dropout. Figure 1 shows a visual summary of the model. For all layers, we use rectifier linear units as activation function. We use a sigmoid activation function on the final layer to generate our prediction for the classification of progression. To mitigate the effects of the highly unbalanced dataset, we oversample the minority class during training.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

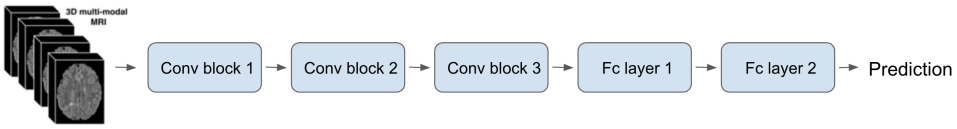

Figure 1: Proposed 3D CNN: Inputs are fed to three repeated 3D convolution blocks, batch normalization and 3D maxpooling. The last layer is flattened and fed to two fully connected layers. We use RELUs as the activation function and dropout with a drop probability of 50 %.

3 Experiments and Results

The ability of the network to prediction clinical progression based on baseline MRI modalities is examined through a series of experiments. We train our model using baseline MRI volumes in the attempt to extract meaningful features leading to progression prediction. Additionally, we combine our CNN with lesion segmentation masks in order to leverage the disease burden information it contains and potentially maximize our model's accuracy. As a comparative baseline, we train a standard 3D CNN consisting of 4 sequential blocks of convolutional, batch normalization, and dropout layers, using both the MRI modalities and the lesion masks as inputs.

3.1 Datasets

Data used for our experiments is drawn from two large proprietary, multi-scanner, multi-center, clinical trial datasets of patients with Relapsing-Remitting MS (RRMS) that was registered in the MNI/ICBM space. Only patients from the placebo arms of both trials were used in the experiments in order to eliminate the drug effects on our analysis of the natural disease course. In addition, only patients who completed the trial were used. One trial consisted of 1330 RRMS patients, out of which 312 patients were on placebo and completed the trial. Two clinical visits at one year apart were available for analysis, resulting in 624 input points with a non-progression/progression split of 582/42. The second dataset consisted of 543 RRMS patients, 153 patients were in the placebo arm and also finished the trial. Three clinical visits, 24 weeks apart, were used in the experiments, resulting in 459 input points with a non-progression/progression split of 398/61. T1-weighted pre- and post- contrast (T1p, T1c), T2-weighted (T2w), Proton Density-weighted (Pdw) and Fluid-attenuated inversion (FLAIR) were used as input modalities for the network. In addition, two lesion masks (T2-weighted and Gadolinium enhanced) were provided as inputs to the network. These masks were provided with the dataset and were obtained previously through a semi-manual process where an automated segmentation algorithm was then corrected by a trained expert reader. Progression labels were assigned to each patient's visit according to the clinical definition progression, measured by the change in the extended disability status scale (EDSS) score over time define by [9] (see Table 1).

Table 1: Requirement for progression given a baseline EDSS score

Definition of Progression	
Baseline EDSS	Criteria
0	An increase of 1.5 or more in EDSS score sustained for 12 weeks or more
0.5 to 5.5	An increase of 1 or more in EDSS score sustained for 12 weeks or more
6.0 and up	An increase of 0.5 or more in EDSS score sustained for 12 weeks or more

3.2 Experiments and Results

The model was first trained using only the five MRI modalities as input. We use the True Positive Rate ($TPR = \frac{TP}{TP+FP}$) and False Positive Rate ($FPR = \frac{TP}{TP+FN}$) receiver operation characteristic (ROC) curve as a performance metric instead of accuracy due to the large class imbalance. The same model is then trained on the MRI inputs with the addition of both T2 lesion and Gad-enhancing lesion masks provided at baseline in order to assess their contribution to the network's performance. We split the dataset into training (75%), validation (15%) and test sets (10%) and perform k-cross

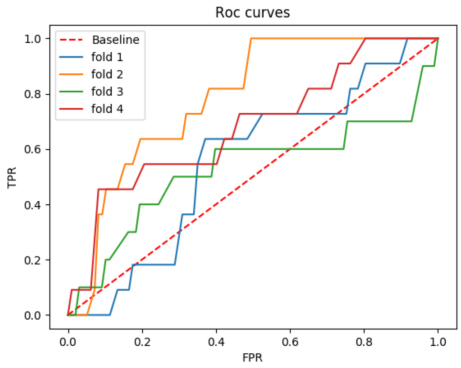

Figure 2: Without lesion labels

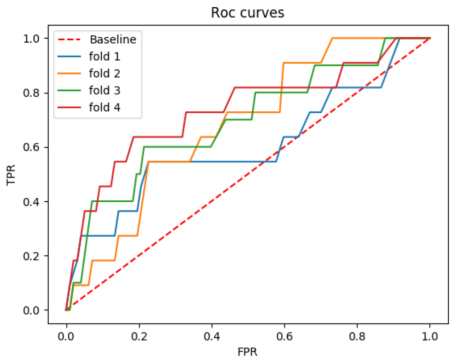

Figure 3: With lesion labels

validation with k=4. As recall and precision are both important for progression prediction, we use the F-score for early stopping.

When using the MRI as input, the network attained an average area under the curve (AUC) of 0.66 ± 0.055. With the addition of T2 lesion and Gad-enhancing lesion masks as inputs, the progression prediction AUC is improved substantially and the variability across folds is reduced. The AUC results for this case was 0.701 ± 0.027. Figure 2 and 3 show the ROC curves for both experiments. Table 2 summarizes our results for all three experiments.

Table 2: Comparison of models' performance

Networks	Number of Parameteres	AUC(std)
Standard 3D CNN	13.5M	0.615 ± 0.053
Proposed 3D CNN	14M	0.66 ± 0.055
Proposed 3D CNN w/ lesion masks	14M	0.701 ± 0.027

Monte Carlo samples of prediction results at test time are acquired, and the model's uncertainty associated with each prediction is estimated based on the MC sample variance estimates. We plot the ROC curves when excluding predictions above certain uncertainty threshold. Figure 4 shows how excluding even a few of the most uncertain predictions can improve the model's overall performance on the remaining predictions.

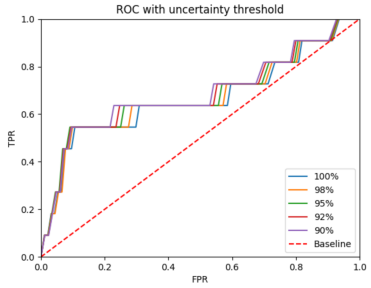

Figure 4: ROC curves at different uncertainty thresholds

4 Discussion and Conclusion

In this paper, we use an end-to-end 3D CNN with parallel convolutional layers for prediction future disability progression in MS patients. We show how by including lesion labels, in addition to multi-modal MRI, the model's performance can be improved. Furthermore, using uncertainty metrics to filter uncertain prediction result in improved operating points on the remaining predictions.

Acknowledgements This work was supported by an award from the International Progressive MS Alliance (PA-1603-08175).

References

[1] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[2] Martin Styner and other. 3d segmentation in the clinic: A grand challenge ii: Ms lesion segmentation. 2008.

[3] Aaron Carass et al. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. *NeuroImage*, 148:77–102, 2017.

[4] Tom Brosch et al. Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE transactions on medical imaging*, 35(5):1229–1239, 2016.

[5] M. Moccia, N. de Stefano, and F. Barkhof. Imaging outcome measures for progressive multiple sclerosis trials. *Multiple Sclerosis Journal*, 23(12):1614–1626, 2017.

[6] J Río et al. MR imaging in monitoring and predicting treatment response in multiple sclerosis. *Neuroimaging Clinics*, 27(2):277–287, 2017.

[7] M. P. Sormani and P. Bruzzi. Mri lesions as a surrogate for relapses in multiple sclerosis: a meta-analysis of randomised trials. *The Lancet Neurology*, 12(7):669–676, 2013.

[8] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.

[9] Rotem Orbach, Zhenming Zhao, Yong-Cheng Wang, Gilmore O'Neill, and Diego Cadavid. Comparison of disease activity in spms and ppms in the context of multicenter clinical trials. *PLoS One*, 7(10):e45409, 2012.