ANALYSIS OF THE INFLUENCE OF DIFFEOMORPHIC NORMALIZATION IN THE PREDICTION OF STABLE VS PROGRESSIVE MCI CONVERSION WITH CONVOLUTIONAL NEURAL NETWORKS

Ubaldo Ramon-Julvez, Monica Hernandez, Elvira Mayordomo and ADNI *

University of Zaragoza

ABSTRACT

We study the effect of the selection of diffeomorphic normalization in the performance of Spasov's deep-learning system for the problem of progressive MCI vs stable MCI discrimination. We considered different degrees of normalization (no, affine and non-rigid normalization) and two diffeomorphic registration methods (ANTS and BL PDE-LDDMM) with different image similarity metrics (SSD, NCC, and lNCC) yielding qualitatively different deformation models and quantitatively different degrees of registration accuracy. BL PDE-LDDMM NCC achieved the best performing accuracy with median values of 89%. Surprisingly, the accuracy of no and affine normalization was also among the highest, indicating that the deep-learning system is powerful enough to learn accurate models for pMCI vs sMCI discrimination without the need for normalization. However, the best sensitivity values were obtained by BL PDE-LDDMM SSD and NCC with median values of 97% and 94% while the sensitivity of the remaining methods stayed under 88%.

Index Terms— CNNs, multi-task learning, diffeomorphic normalization, Alzheimer's, pMCI vs sMCI

1. INTRODUCTION

Alzheimer's disease is one of the most common neurodegenerative diseases [1]. This disease arises for still unknown reasons and progresses relentlessly towards a devastating cognitive condition. Research against neurodegeneration is approached in two antagonistic ways: while some researchers work to develop treatments to slow or reverse the loss of cognitive abilities, others are working on developing ways to diagnose the cause of dementia as early as possible. Although our knowledge in Alzheimer's disease has considerably increased in the last two decades, the percentage of patients that can benefit from current knowledge is limited compared with the estimated population at risk. Many individuals can be diagnosed with probable Alzheimer's when the disease has started to manifest its symptoms, losing valuable time for preventive or early treatment.

Especially urgent is the case of subjects with progressive mild cognitive impairment (pMCI). They are patients that have started suffering a gentle cognitive decay that will accelerate until the disease develops over the next three years. We need predictive biomarkers and accurate diagnostic tools of the rates of cognitive decline for those individuals who exhibit preclinical or prodromal symptoms. The quest for significant biomarkers and the design of predictive diagnostic tools go through manipulation of huge amounts of data for each patient. Therefore, clinical practice must ally with the computational tools that facilitate the processing of such amount of data.

This work focuses on the problem of discriminating progressive vs stable MCI individuals (pMCI vs sMCI) with computational tools. This problem has been previously approached using different combinations of biomarkers with conventional machine learning methods [2, 3, 4, 5, 6, 7]. From them, the most discriminative power has been achieved combining demographic and cognitive biomarkers, APOE status, and anatomical MRI with accuracy ranging between 80 - 85% and sensitivity ranging between 80 - 89%.

The deep-learning explosion has reached the area of computer-aided diagnosis with very powerful tools for the discrimination of healthy vs diseased individuals and the prediction of patient condition [8]. The problem of pMCI vs sMCI has been recently approached using different deep-learning systems in [9, 10]. The convolutional neural network proposed in Spasov et al. [10] outperformed conventional machine learning methods. The fundamental problem faced by this method was the scarcity of data, typical in medical applications. The authors approached the problem combining pMCI vs sMCI with AD vs HC problems in a multi-task learning framework. In addition, they proposed a parameter reduction replacing 3D convolutions by separable convolutions in the middle hidden layers of the network. The method used the combination of the best performing biomarkers in support vector machine (SVM) systems. MRI was selected from imaging due to its less invasive nature. The MRI images were normalized to a common reference system using affine

*Data used in preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). As such, the investigators within the ADNI just contributed to the design and implementation of ADNI and/or provided data. We would like to thank Mr. Simeon Spasov for expert advice and Profs. Toschi and Lio for feedback. Work supported with grant TIN2016-80347-R.

plus diffeomorphic registration, as it is customary in morphometric methods such as Voxel, Tensor, or Deformation Based Morphometry (V/T/DBM) [11]. The authors compared the performance of the network combining clinical and genetic data with typical inputs in V/T/DBM in the input stream, such as the aligned MRIs, the Jacobian determinants, or both.

The authors also analyzed the robustness of the network to variations in the selection of the reference system and the location of the anatomical information, which have shown to be crucial in V/T/DBM studies [12]. Thus, the performance of the network was compared between the normalization to a data-specific atlas and the MNI152 atlas. In addition, the authors compared the use of skull-stripped images in the input stream with masked images in the areas of most interest in Alzheimer's disease. In both cases, the network was able to identify the most relevant features for the pMCI vs sMCI problem showing a comparable accuracy. Although the quality of the image registration is relevant in the outcome of V/T/DBM studies, the robustness of the network to variations of the quality of diffeomorphic normalization was not analyzed.

In this work, we study the effect of the selection of the diffeomorphic normalization in the performance of Spasov et al. deep-learning system for the problem of pMCI vs sMCI. We have compared different degrees of normalization: no normalization, affine normalization, and affine plus diffeomorphic normalization to the MNI152 atlas. We have included in our study the following diffeomorphic registration methods: 1) ANTS diffeomorphic registration [13] with Sum of Squares Differences (SSD) and local Normalized Cross-Correlation (lNCC), and 2) BL PDE-LDDMM [14] with SSD, NCC, and lNCC.

2. DESCRIPTION OF THE CNN SYSTEM

2.1. Network architecture

We reproduced the network architecture proposed in [10]. The network combined 3D convolutional layers (3D-CNN) with separable convolutional layers (sCNN) and fully connected layers (FC) in nine hidden layers. For separable convolution, we used the Keras layer based on TensorFlow provided by the authors in https://github.com/simeon-spasov/MCI. The system was fed with three different input streams that can be activated independently. We combined demographic, cognitive, and genetic data into a single input stream with clinical information. The system was fed with the MRI and Jacobian information separately, and concatenated before the separable convolutions. The clinical stream was concatenated with the output of the image-based FC layer in the deepest layer of the system. No data augmentation procedures were used in the system. In order to tackle the problem of data scarcity, the network was used in two different classification problems. The first one was Alzheimer's Disease (AD) vs Healthy Control (HC) discrimination. The second one was our target pMCI vs sMCI discrimination. Thus, the data available for training, validation, and testing increased, and the confidence of the network in the AD vs HC problem helps to improve the performance in the more difficult pMCI vs sMCI task. Fine details of the architecture and the implementation can be found in [15, 10].

2.2. Datasets and preprocessing

The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI is to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The authors from [10] kindly provided us the IDs of the individuals used in their study, facilitating reproducibility. From them, 191 subjects belonged to AD, 181 to HC, 179 to pMCI, and 227 to sMCI. It should be noticed that we used a total of 778 individuals. This number differs slightly from the 785 individuals used in [10] due to corrupted images and lack of clinical data. From the seven missing individuals, two belonged to pMCI class and just one to sMCI.

The T1 images from the baseline visit were used in our study. The images were preprocessed using the standard pipeline proposed in [16] for neuroimaging studies. The images were reoriented and affinely normalized to the Montreal Neurological T1 Template (MNI152) atlas using FSL package (https://fsl.fmrib.ox.ac.uk). The bias field was corrected using the N4 method available in ANTS software (http://stnava.github.io/ANTs). The images were skull stripped using Robbex (https://www.nitrc.org/ projects/robex). Finally, affine plus diffeomorphic normalization was performed towards the MNI152 atlas.

2.3. Methods considered for diffeomorphic normalization

2.3.1. Advanced Normalization ToolS

Advanced Normalization ToolS (ANTS) is a comprehensive toolkit intended for rigid and diffeomorphic normalization (http://stnava.github.io/ANTs/). The primary diffeomorphic registration method is based on Avants et al. SyN registration [13]. The most relevant features of the method are the symmetric steady approach to Large Deformation Diffeomorphic Metric Mapping (LDDMM), the use of lNCC similarity metric, and the use of a multiresolution approach with kernel regularization of varying width. This method has won the main unbiased and international competitions in image registration in the last decade. ANTS is provided as open-source C++ software, closely depending on the Insight

Toolkit (http://www.itk.org). The implementation is only runnable on the CPU.

In this study, we consider two different image similarity metrics: SSD and lNCC. For the same transformation characterization, regularization, and optimization, the use of lNCC image similarity usually outperforms SSD in typical non-rigid registration evaluation frameworks [17]. Therefore, the presumable worse registration accuracy of ANTS-SSD vs ANTS-lNCC can be used to analyze the effect of the selection of the diffeomorphic normalization in the performance of our system.

2.3.2. Band Limited PDE-constrained LDDMM

Band Limited PDE-constrained LDDMM (BL PDE-LDDMM) is the best performing method proposed in [14]. BL PDE-LDDMM belongs to the family of physically meaningful diffeomorphic registration methods. These methods model the problem using a PDE-constrained variational formulation that provides the versatility to impose different physical models to the computed transformations by just adding the PDEs associated with the problem as hard constraints. Numerical optimization is approached in the form of inexact Newton-Krylov methods which has shown an excellent numerical accuracy and an extraordinarily fast convergence rate in this framework. The band-limited parameterization is one strategy successfully circumventing the huge computational complexity of PDE-LDDMM methods without much loss in the accuracy. Indeed, the method is implemented in the GPU.

In this study we consider three different image similarity metrics: SSD, NCC, and lNCC. As happened with ANTS, lNCC image similarity outperforms SSD in typical non-rigid registration evaluation frameworks [17]. NCC and lNCC perform similarly.

3. RESULTS

In this experimental section, we study the effect of the selection of the diffeomorphic normalization in the performance of our developed system for the problem of pMCI vs sMCI [1]. We consider as baseline the results obtained in Spasov et al. [10], which is the best performing method in the state-of-the-art to the date. Table 1 shows the median of the area under the curve (AUC), the accuracy (ACC), the sensitivity (SEN), and the specificity (SPE) obtained from independent test sets across the 10-fold experiments for three different anatomical input streams: the MRI images, the Jacobian determinant of the diffeomorphic transformations, and the MRI images plus Jacobian determinant. AUC was calculated by performing Receiver operating characteristic (ROC) on each fold, and all accuracy,sensitivity and specificity metrics are reported at the optimal operating point of the ROC curve obtained via the Youden's J statistic. Table 2 shows the results obtained in [10] for the same input streams and diffeomorphic normalization using ANTS-lNCC with respect to MNI152 atlas and a data-specific atlas.

[1] It should be noticed that our system obtained 100% scores in the AD vs HC problem.

Table 1. Median of the performance metrics on the pMCI vs sMCI problem using our deep-learning implementation with the variants considered for diffeomorphic normalization.

Normalization	Anat. data	AUC	ACC	SEN	SPE
Unregistered	MRI	0.92	88	81	88
Affine	MRI	**0.94**	**89**	88	**94**
ANTS-SSD	MRI	0.90	86	84	88
ANTS-lNCC	MRI	0.90	84	88	81
BL PDE-LDDMM SSD	MRI	0.90	88	**97**	81
BL PDE-LDDMM NCC	MRI	**0.94**	**89**	94	91
BL PDE-LDDMM lNCC	MRI	0.89	84	84	91
ANTS-SSD	Jac	0.86	83	84	84
ANTS-lNCC	Jac	0.83	80	72	88
BL PDE-LDDMM SSD	Jac	0.87	86	84	84
BL PDE-LDDMM NCC	Jac	0.90	88	88	81
BL PDE-LDDMM lNCC	Jac	0.91	86	91	81
ANTS-SSD	MRI + Jac	0.89	84	88	81
ANTS-lNCC	MRI + Jac	0.88	83	88	84
BL PDE-LDDMM SSD	MRI + Jac	0.90	84	88	81
BL PDE-LDDMM NCC	MRI + Jac	0.91	88	88	88
BL PDE-LDDMM lNCC	MRI + Jac	**0.94**	**89**	91	88

Table 2. Median of the performance metrics on the pMCI vs sMCI problem obtained in Spasov et al. [10].

Normalization	Atlas	Anat. data	AUC	ACC	SEN	SPE
ANTS-lNCC	MNI152	MRI	0.91	85	82	87
ANTS-lNCC	MNI152	Jac	0.88	82	82	81
ANTS-lNCC	MNI152	MRI + Jac	0.89	83	77	86
ANTS-lNCC	data-specific	MRI	0.92	86	87.5	84
ANTS-lNCC	data-specific	Jac	0.87	83	84	78
ANTS-lNCC	data-specific	MRI + Jac	0.91	83	87	81

The best performing anatomical input stream was compound by the MRI images for almost all the normalizations. This result corroborates the results obtained in [10]. For the MRI input stream, the accuracy obtained by our system overpassed the 86% obtained by the baseline system except in the lNCC cases. Interestingly, the use of no diffeomorphic normalization and affine normalization achieved an accuracy of 88 and 89%, respectively, reaching the performance of the best diffeomorphic normalization method BL PDE-LDDMM NCC. For the same image similarity metric, BL PDE-LDDMM normalization overpassed ANTS. For BL PDE-LDDMM the NCC image similarity metric outperformed SSD and lNCC. One tailed Mann-Whitney U test did not report statistical significance for BL PDE-LDDMM NCC (p-values between 0.18 and 0.63). For affine registration statistical significance was found for ANTS-SSD, ANTS-lNCC, and BL PDE-LDDMM lNCC (p = 0.07, 0.05, and 0.07 respectively). For the Jacobian and MRI plus Jacobian input streams, BL PDE-LDDMM with NCC and lNCC image similarity were the best performing methods, respectively.

For the MRI input stream, the sensitivity of our system

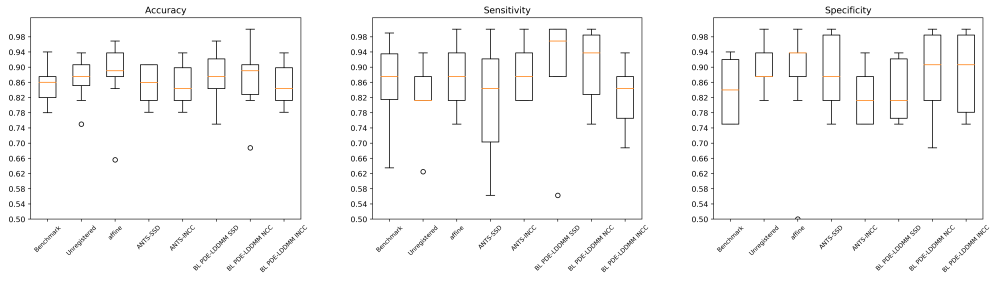

Fig. 1. Box plots for the ACC, SEN, and SPE obtained over the 10-folds in the case of MRI anatomical input stream. The benchmark reproduces the results of Figure 6 in [10] for the MRI and clinical input stream.

was under 90% for no normalization, affine, ANTS, and BL PDE-LDDMM with lNCC normalization. The sensitivity of BL PDE-LDDMM with SSD and NCC reached values ranging from 94 to 97%. Statistical significance was found with respect to no, ANTS-SSD, ANTS-lNCC, and BL PDE-LDDMM lNCC (p= 0.01, 0.05, 0.09, and 0.01, respectively). The best specificity values were obtained by no and affine normalization, with a median of 94% sMCI subjects correctly identified. Statistical significance was found with respect to no, ANTS-lNCC, and BL PDE-LDDMM SSD (p= 0.00 and 0.03). For the Jacobian and MRI plus Jacobian input streams, it is remarkable the sensitivity of BL PDE-LDDNN lNCC method with 91%.

Figure 1 shows in the shape of box-and-whisker the performance values over the 10-folds obtained by the different diffeomorphic normalization in the case of MRI input stream.

4. DISCUSSION AND CONCLUSIONS

In this work, we have studied the effect of the selection of the diffeomorphic normalization in the performance of Spasov et al. deep-learning system for the problem of pMCI vs sMCI identification. We have compared different degrees of normalization ranging from no normalization to affine plus diffeomorphic normalization As anatomical input streams we have used the aligned MRIs, Jacobian determinants, or both.

We have studied the performance of two different families of diffeomorphic registration methods: celebrated ANTS diffeomorphic registration [13] and an efficient version of physically meaningful LDDMM, BL PDE-LDDMM [14]. We have implemented BL PDE-LDDMM with different image similarity metrics providing an interesting variability of accuracies in traditional registration evaluation studies [17].

We were able to roughly reproduce the results presented in [10] despite the complexity of the pipeline (see Fig. 1). In addition, we corroborated the robustness of the network to variations in the selection of the reference system and the use of MRI as the best performing anatomical input stream.

The best performing diffeomorphic normalization was obtained with BL PDE-LDDMM with NCC image similarity. Surprisingly, the accuracy achieved by the no and affine normalization was similar to the accuracy achieved by BL PDE-LDDMM NCC. This indicates that the deep-learning system is powerful enough to learn accurate models for pMCI vs sMCI discrimination without the need for normalization. In contrast, less powerful systems need the non-rigid alignment of the images in order to discriminate the MCI patients from the anatomical features [4, 6].

Regarding the sensitivity of the system, the best values were obtained by BL PDE-LDDMM with the global image similarity metrics (SSD and NCC). It seems that non-rigid normalization with a physically meaningful deformation model may be translated into values of sensitivity better than ANTS deformation model. In addition, it seems that the use of no or affine normalization may be translated into a worse capacity to predict pMCI individuals. For ANTS diffeomorphic registration, the better registration accuracy of lNCC image similarity is translated into a better sensitivity. However, for BL PDE-LDDMM the worst sensitivity was precisely obtained with the lNCC metric. Paradoxically, the sensitivity with the lNCC metric and the Jacobian involved input streams was the highest. We need further experimentation to find out why is BL PDE-LDDMM lNCC decreasing the prediction capacity of pMCI individuals for the MRI input streams.

The best specificity results were achieved by the no and affine normalization, followed by BL PDE-LDDMM with NCC and lNCC metrics. We believe that since sMCI group includes different classes of cognitively impaired individuals, the diffeomorphic alignment may not be contributing to better identification of the stable individuals.

These results position BL PDE-LDDMM over ANTS for diffeomorphic normalization in Spasov et al. system for the problem of pMCI vs sMCI and leave open the question of whether performing no or affine normalization would be enough for the system to obtain overall satisfactory results. In future work, we will extend our comparison to the whole ADNI project and we will evaluate the performance of the system with other families of diffeomorphic registration methods with underlying interesting deformation models.

5. REFERENCES

[1] Alzheimer's association, "Alzheimer's facts and figures," https://www.alz.org/media/documents/alzheimers-facts-and-figures-2019-r.pdf, 2019.

[2] E. Moradi, A. Pepe, C Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *Neuroimage*, vol. 104, pp. 398 – 412, 2015.

[3] I. O. Korolev, L. L. Symonds, and A. C. Bozoki, "Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical MRI and plasma biomarkers via probabilistic pattern classification," *PLoS One*, vol. 11, 2016.

[4] L. Beheshti, H. Demirel, and H. Matsuda, "Classification of Alzheimer's disease and prediction of mild cognitive impairment to Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Comput. Biol. Med.*, vol. 83, pp. 109 – 119, 2017.

[5] K. Liu, K. Chen, and L. Guo, "Prediction of mild cognitive impairment conversion using a combination of independent component analysis and the Cox model," *Front. Hum. Neurosci.*, vol. 11, 2017.

[6] T. Tong et al. and ADNI, "A novel grading biomarker for the prediction of conversion from mild cognitive imparirment to Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 64, pp. 155 – 165, 2017.

[7] M. Grassi, N. Rouleaux, D. Caldirola, D. Loewenstein, K. Schruers, G. Perna, M. Dumontier, and ADNI, "A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures," *Front. Neurol.*, vol. 10, pp. 756, 2019.

[8] L. Lu, Y. Zheng, G. Carneiro, and L. Yang, "Deep learning and convolutional neural networks for medical image computing," *Springer*, 2017.

[9] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multiscale deep nerual network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 46, pp. 26 – 34, 2018.

[10] S. E. Spasov, L. Passamonti, A. Duggento, P. Lio, N. Toschi, and ADNI, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *Neuroimage*, vol. 189, pp. 276 – 287, 2019.

[11] X. Hua, A. D. Leow, N. Parikshak, S. Lee, M. C. Chiang, A. W. Toga, C. R. Jack, M. W. Weiner, P. M. Thompson, and ADNI, "Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: an MRI study of 676 AD, MCI, and normal subjects," *Neuroimage*, vol. 43(3), pp. 458 – 469, 2008.

[12] J. Koikkalainen, J. Lotjonen, L. Thurfjell, D. Rueckert, G. Waldemar, H. Soininen, and ADNI, "Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease," *Neuroimage*, vol. 56(3), pp. 1134 – 1144, 2011.

[13] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric dieomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, pp. 26 – 41, 2008.

[14] M. Hernandez, "A comparative study of different variants of Newton-Krylov PDE-constrained Stokes-LDDMM parameterized in the space of band-limited vector fields," vol. 12(2), pp. 1038 – 1070, 2019.

[15] S. E. Spasov, L. Passamonti, A. Duggento, P. Lio, N. Toschi, and ADNI, "A multi-modal convolutional neural network framework for the prediction of Alzheimer's disease," 2018.

[16] M. HadjHamou, M. Lorenzi, N. Ayache, and X. Pennec, "Longitudinal analysis of image time series with diffeomorphic deformations: a computational framework based on stationary velocity fields," *Frontiers in Neuroscience*, 2016.

[17] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M. C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *Neuroimage*, vol. 46(3), pp. 786–802, 2009.