Using diffusion MRI for classification and prediction of Alzheimer's Disease: a reproducible study

Junhao Wen, Jorge Samper-Gonzalez, Simona Bottani, Alexandre Routier, Ninon Burgos, Thomas Jacquemont, Sabrina Fontanella, Stanley Durrleman, Anne Bertrand, Olivier Colliot

Using diffusion MRI for classification and prediction of Alzheimer's Disease: a reproducible study

Junhao Wen[1,2], Jorge Samper-González[1,2], Simona Bottani[1,2], Alexandre Routier[1,3], Ninon Burgos[1,2], Thomas Jacquemont[1,2], Sabrina Fontanella[1,2], Stanley Durrleman[1,2], Anne Bertrand[1,4,5], Olivier Colliot[1,4,6]

[1]Inria Paris, Aramis project-team, Paris, France
[2]Sorbonne Université, Inserm, CNRS, Institut du Cerveau et la Moëlle (ICM), Paris, France
[3]Sorbonne Université, Inserm, CNRS, Institut du Cerveau et la Moëlle (ICM), FrontLab, Paris, France
[4]Sorbonne Université, Inserm, CNRS, Institut du Cerveau et la Moëlle (ICM), AP-HP, Paris, France
[5]AP-HP, Saint-Antoine Hospital, Department of Radiology, Paris, France
[6]AP-HP, Departments of Neuroradiology and Neurology, Pitié-Salpêtrière Hospital, Paris, France

Background

Several studies using machine learning have recently looked at the potential of diffusion tensor imaging (DTI) for Alzheimer's disease (AD) classification (O'Dwyer et al., 2012; Dyrba et al., 2013; Maggipinto et al., 2017). However, classification accuracies are not directly comparable across studies because of differences in subject selection, image processing, feature extraction and selection, and classification algorithms. Samper-Gonzalez et al. (2017) proposed a reproducible framework for automatic classification of AD from T1 MRI and PET data. Here, we extend this framework for DTI-based classification using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. This work aims to facilitate replication of classification experiments based on ADNI and also to compare the classification performances with different DTI-based features.

Methods

The framework is composed of the following components. Tools were implemented to automatically convert original ADNI diffusion MRI into the BIDS format. Standardized artifact correction procedures were designed. The DTI model was then fitted generating FA and MD maps. These maps were nonlinearly registered to MNI space and masked using the WM, GM or GM+WM binarized maps. Classification was performed using a linear support vector machine (SVM). A repeated holdout cross-validation (250 runs) with a 10-fold inner grid search for hyperparameter optimization was performed. Additionally, optimal margin hyperplane coefficient maps were reported to characterize potential anatomical patterns (Cuingnet et al, 2013). We demonstrate the use of the framework with 46 AD patients, 46 cognitively normal (CN) subjects, 97 mild cognitive impairment (MCI) subjects, including 54 stable MCI (sMCI) and 24 progressive MCI (pMCI) at 36 months. Additionally, a random undersampling technique with 10 repetitions was used to assess the impact of imbalanced data (e.g 54 sMCI vs 24 pMCI) on the classification performance.

Results

Classification results are shown in tables 1 to 4. We found high accuracies for AD vs CN classification but lower accuracies for classifications involving MCI. Balanced and imbalanced classifications gave comparable accuracies. The hippocampus and medial temporal cortex are the main discriminative regions using WM+GM-MD features (Fig 1).

Conclusions

We presented a framework that enables the comparability and reproducibility of DTI-based AD classification. Classification results are in line with the state of the art. The low accuracies obtained in discriminating pMCI from sMCI or CN are potentially caused by the small data samples, rather than by imbalanced data. The code will be made publicly available at https://github.com/aramis-lab/AD-ML.

References

Samper-Gonzalez, J., Burgos, N., Fontanella, S., Bertin, H., Habert, M.O., Durrleman, S., Evgeniou, T., Colliot, O. and Alzheimer's Disease Neuroimaging Initiative, 2017, September. Yet Another ADNI Machine Learning Paper? Paving The Way Towards Fully-reproducible Research on Classification of Alzheimer's Disease. In International Workshop on Machine Learning in Medical Imaging (pp. 53-60). Springer, Cham.

Cuingnet, R., Glaunès, J.A., Chupin, M., Benali, H. and Colliot, O., 2013. Spatial and anatomical regularization of SVM: a general framework for neuroimaging data. IEEE transactions on pattern analysis and machine intelligence, 35(3), pp.682-696.

Task	Feature	AUC	Balanced Accuracy	Sensitivity	Specificity
46 CN vs 46 AD	WM-FA	0.82	0.76	0.77	0.74
	GM-FA	0.77	0.71	0.70	0.74
	WM+GM-FA	0.82	0.77	0.76	0.78
	WM-MD	0.82	0.74	0.63	0.86
	GM-MD	0.83	0.76	0.68	0.83
	WM+GM-MD	0.82	0.75	0.67	0.82

Table 1 Classification results obtained with DTI-based features using a linear SVM classifier when differentiating CN (n=46) from AD (n=46) subjects. The balanced accuracy, sensitivity and specificity were averaged over 250 runs. AUC: area under the receiver operating characteristic (ROC) curve.

Task	Feature	AUC	Balanced Accuracy	Sensitivity	Specificity
46 CN vs 97 MCI	WM-FA	0.65	0.60	0.64	0.56
	GM-FA	0.68	0.61	0.74	0.48
	WM+GM-FA	0.68	0.61	0.74	0.47
	WM-MD	0.68	0.62	0.53	0.71
	GM-MD	0.61	0.61	0.57	0.65
	WM+GM-MD	0.62	0.62	0.58	0.67
46 CN vs 46 MCI	WM-FA	0.61	0.60	0.54	0.66
	GM-FA	0.60	0.59	0.52	0.67
	WM+GM-FA	0.67	0.63	0.57	0.69
	WM-MD	0.54	0.59	0.39	0.78
	GM-MD	0.50	0.59	0.49	0.68
	WM+GM-MD	0.48	0.57	0.45	0.68

Table 2 Classification results obtained with DTI-based features using a linear SVM classifier when differentiating CN (n=46) from MCI (n=97) subjects. The balanced accuracy, sensitivity and specificity were averaged over 250 runs for no sampling classifications, and 10*250 runs for undersampling classifications. AUC: area under the receiver operating characteristic (ROC) curve.

Task	Feature	AUC	Balanced Accuracy	Sensitivity	Specificity
46 CN vs 24 pMCI	WM-FA	0.60	0.58	0.85	0.32
	GM-FA	0.77	0.65	0.87	0.43
	WM+GM-FA	0.78	0.66	0.88	0.44
	WM-MD	0.56	0.65	0.84	0.47
	GM-MD	0.55	0.61	0.80	0.42
	WM+GM-MD	0.54	0.61	0.83	0.4
24 CN vs 24 pMCI	WM-FA	0.59	0.61	0.70	0.52
	GM-FA	0.70	0.66	0.71	0.60
	WM+GM-FA	0.76	0.69	0.75	0.63
	WM-MD	0.33	0.63	0.88	0.37
	GM-MD	0.59	0.65	0.72	0.58
	WM+GM-MD	0.62	0.66	0.77	0.55

Table 3 Classification results obtained with DTI-based features using a linear SVM classifier when differentiating CN (n=46) from pMCI (n=24) subjects. The balanced accuracy, sensitivity and specificity were averaged over 250 runs for no sampling classifications, and 10*250 runs for undersampling classifications. AUC: area under the receiver operating characteristic (ROC) curve.

Task	Feature	AUC	Balanced Accuracy	Sensitivity	Specificity
54 sMCI vs 24 pMCI	WM-FA	0.49	0.48	0.32	0.64
	GM-FA	0.48	0.52	0.25	0.78
	WM+GM-FA	0.49	0.52	0.22	0.81
	WM-MD	0.49	0.49	0.66	0.32
	GM-MD	0.44	0.51	0.52	0.50
	WM+GM-MD	0.45	0.50	0.65	0.35
24 sMCI vs 24 pMCI	WM-FA	0.50	0.47	0.47	0.47
	GM-FA	0.50	0.57	0.54	0.59
	WM+GM-FA	0.48	0.56	0.56	0.56
	WM-MD	0.54	0.48	0.56	0.41
	GM-MD	0.45	0.53	0.41	0.65
	WM+GM-MD	0.44	0.54	0.43	0.65

Table 4 Classification results obtained with DTI-based features using a linear SVM classifier when differentiating sMCI (n=54) from pMCI (n=24) subjects. The balanced accuracy, sensitivity and specificity were averaged over 250 runs for no sampling classifications, and 10*250 runs for undersampling classifications. AUC: area under the receiver operating characteristic (ROC) curve.
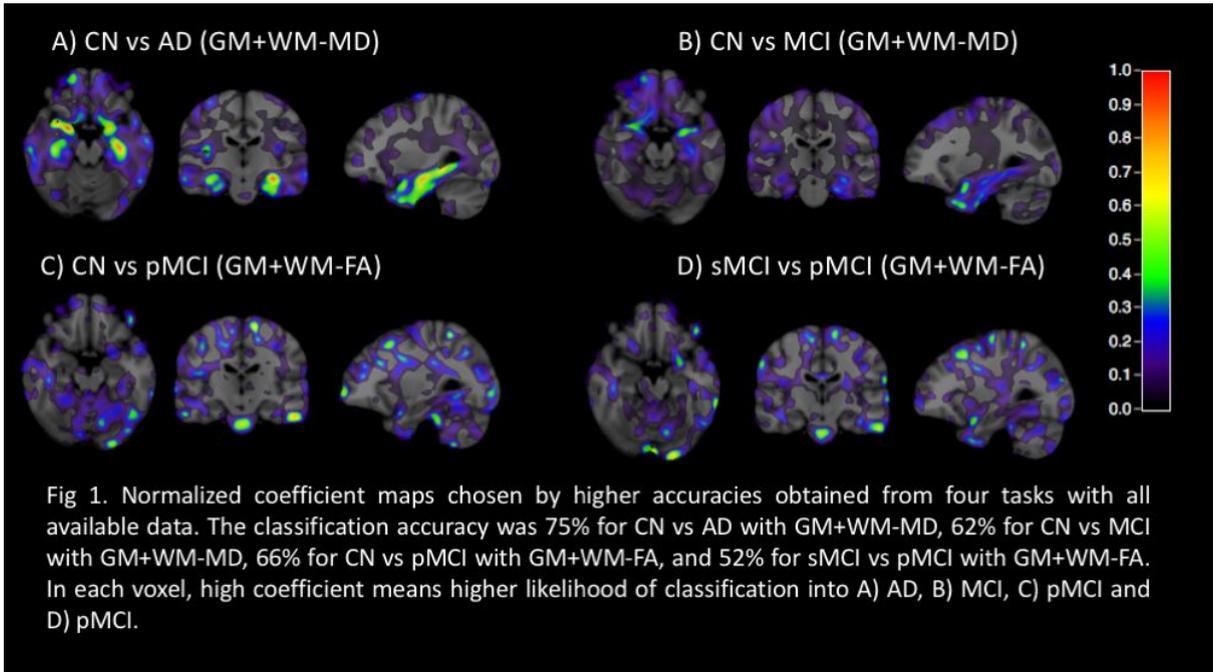
Fig 1. Normalized coefficient maps chosen by higher accuracies obtained from four tasks with all available data. The classification accuracy was 75% for CN vs AD with GM+WM-MD, 62% for CN vs MCI with GM+WM-MD, 66% for CN vs pMCI with GM+WM-FA, and 52% for sMCI vs pMCI with GM+WM-FA. In each voxel, high coefficient means higher likelihood of classification into A) AD, B) MCI, C) pMCI and D) pMCI.