Influence of acquisition parameters on morphometric values obtained from structural magnetic resonance images

Award:	Certificate of Merit
Poster No.:	C-2108
Congress:	ECR 2019
Type:	Scientific Exhibit
Authors:	I. Evangelista, C. L. Galimberti, G. Pascariello, J. C. Gomez, A. L. Rodríguez Musso, P. Donnelly-Kehoe; Rosario/AR
Keywords:	Artificial Intelligence, Anatomy, Neuroradiology brain, MR, Segmentation, Computer Applications-General, Computer Applications-Detection, diagnosis, Dementia
DOI:	10.26044/ecr2019/C-2108

Aims and objectives

Structural MRI (sMRI) of the brain in the clinical environment usually consist in a qualitative analysis, in which the neuroradiologist detects deviations or anomalies by visual inspection. A quantitative analysis can complement the specialist's opinion, allowing a comparative study among subjects. However, to perform a reliable analysis, it is necessary to consider each measure's distribution in the healthy population. In this work we used a large sample from several public databases to study the influence of acquisition parameters on the measurement of neuromorphometric features computed with Freesurfer. Then, we applied them to study Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) from a machine learning perspective.

Quantitative approaches to sMRI are ideal for studying changes in brain structures due to neurodegenerative diseases. For the purpose of developing a system that can help to assess structural abnormalities for individualized medicine, 2700 T1-MRI images were collected from public databases. These data correspond to healthy controls (HC) and both AD and MCI patients. The T1-MRI were acquired under different protocols according to the procedures of the center where participants were scanned. Figure 1 shows the age distribution of participants included in this analysis.

From each T1-MRI a set of quantitative features characterising each brain's anatomy was obtained using FreeSurfer v6.0.

While open databases allow to collect a vast amount of images, combining data from different sources is not straightforward. Different acquisition parameters may bias Freesurfer's morphologic measurements [1]. Indeed, T1-MRI brain images obtained using different field strengths were classified with 88% of accuracy. According to our analysis, differences introduced by field strength were more relevant than those caused by other factors, including the manufacturer, the machine model and the voxel volume.

The possibility of having images obtained using different acquisition parameters is important in the interest of achieving generalization. However, in order to attain a system that can prove robust against changes in acquisition parameters it is required to develop a technique that can minimize the differences caused by them. Therefore, different strategies were explored with the objective of harmonizing data from different sources. The effectiveness of the proposed solutions was analysed from a machine learning perspective. Results showed that it is necessary to take into account brain measures variability with age so as to properly mitigate the influence of the acquisition parameters. Harmonization between field strength groups makes it easier to compare images from different sources, albeit there are a number of variables that cannot be used indistinctly.

Thus, an analysis of feature importance was conducted in order to determine which variables are not susceptible to the compatibilization process.

What is expected from a harmonization process is that it does not affect the the capacity of classifying diseases using the data. Morphological measurements after applying the compatibilization technique proved as effective as the original data to classify AD and MCI.

Images for this section:

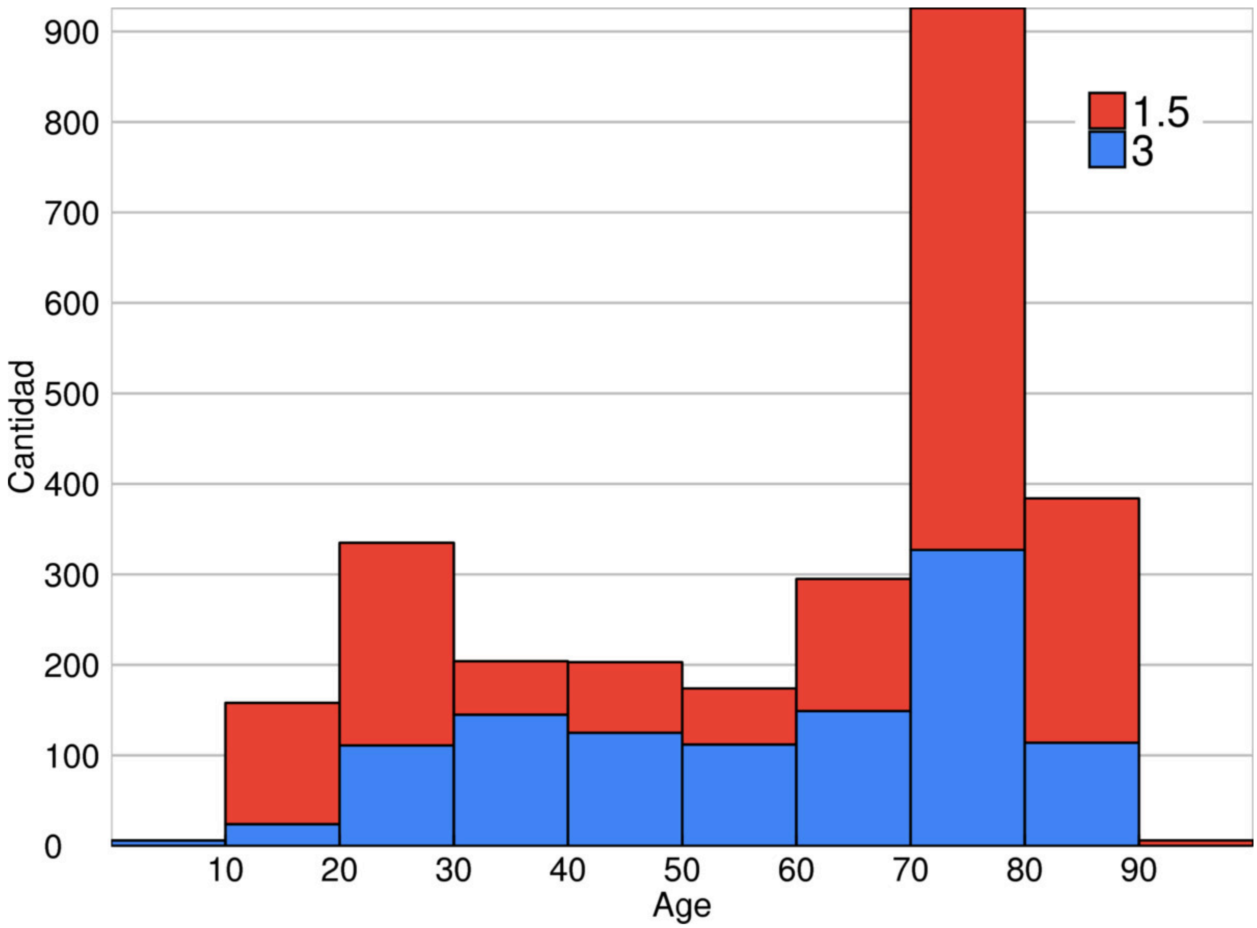

Fig. 1: Age distribution of people included in the present analysis

Methods and materials

FreeSurfer [1] is a software package that can enhance the visual information in resonance images by means of performing structural measurements. Using FreeSurfer v6.0, 2700 MRIs were processed and 272 morphometric features that characterize each brain's cortical and subcortical anatomy were obtained. An example of FreeSurfer segmentation is shown in Figure 2, where each colour represents a cortical or subcortical structure. From the segmentation of these structures, thickness, area and volume values can be measured.

The effects of the different acquisition parameters on the morphological variables were tested using dimensionality reduction approaches and supervised machine learning techniques (random forests). We trained random forest classifiers implemented in Python's Scikit-Learn [2] using the neuromophometric measurements of HC as features. The classification objective was to determine certain acquisition parameter, for example the field strength (1.5T vs 3T) or manufacturer (Simens, General Electric or Philips); that is to say, a classifier was supposed to determine an acquisition parameter based solely on structural information acquired using FreeSurfer. Classification accuracy was taken as an indicator of the extent to which brain measures are different between classes. Since a random forest classifier outputs classification probabilities, these values were also taken as an indicator of the degree of compatibilization achieved. In simpler words, the more certain the classifier was about the output class, the stronger the differences between the classes.

Different solutions were explored to achieve compatibilization: z-score using healthy controls of each group to determine the mean and standard deviation; w-score using healthy controls of each group to determine the linear regression; z-score taking into account age (greater or lower than certain threshold).

Moreover, given the natural aging process of the brain, it was considered essential that the compatibilization technique took into account the changes that brain structures experience with age [3]. We ultimately proposed binning the age and calculating a z-score for each group using healthy controls to determine the mean and standard deviation. Figure 3 shows how the morphometric values change with age for a randomly chosen structure (medial orbitofrontal cortex), and the means calculated for 1.5 T and 3 T for each 10-year bin.

Further, we performed a progressive feature elimination (PFE) [4] in order to determine which variables were important for distinguishing data acquired with different parameters and to analyse the presence of features that were invulnerable to the compatibilization process, i.e. whether there were features that contained information for separating the classes in spite of the proposed compatibilization method.

Finally, in order to corroborate that the data classification power was not modified after the compatibilization, we analyzed how the data was classified depending on the diagnosis: HC, MCI and AZ. Random forest classifiers were developed using one third of the data to perform feature selection and a 10-fold cross validation to evaluate the model.

Images for this section:

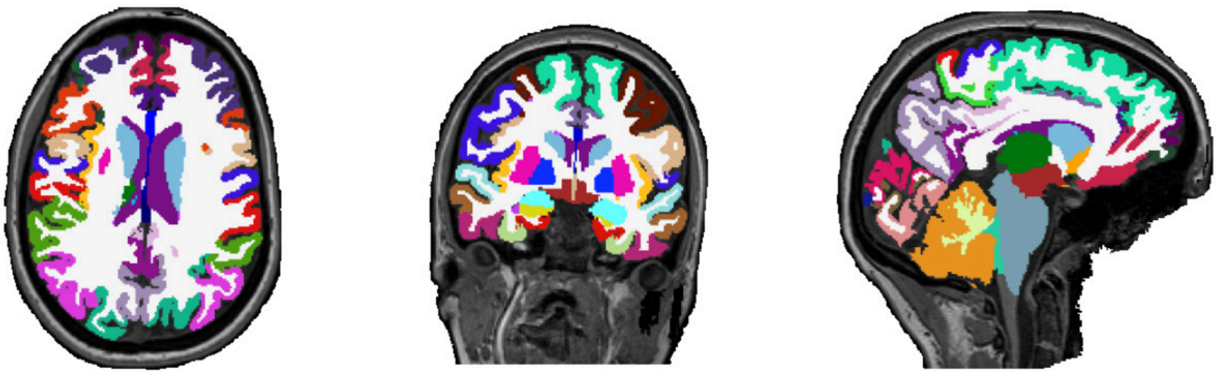

Fig. 2: Example of a brain segmentation using FreeSurfer.

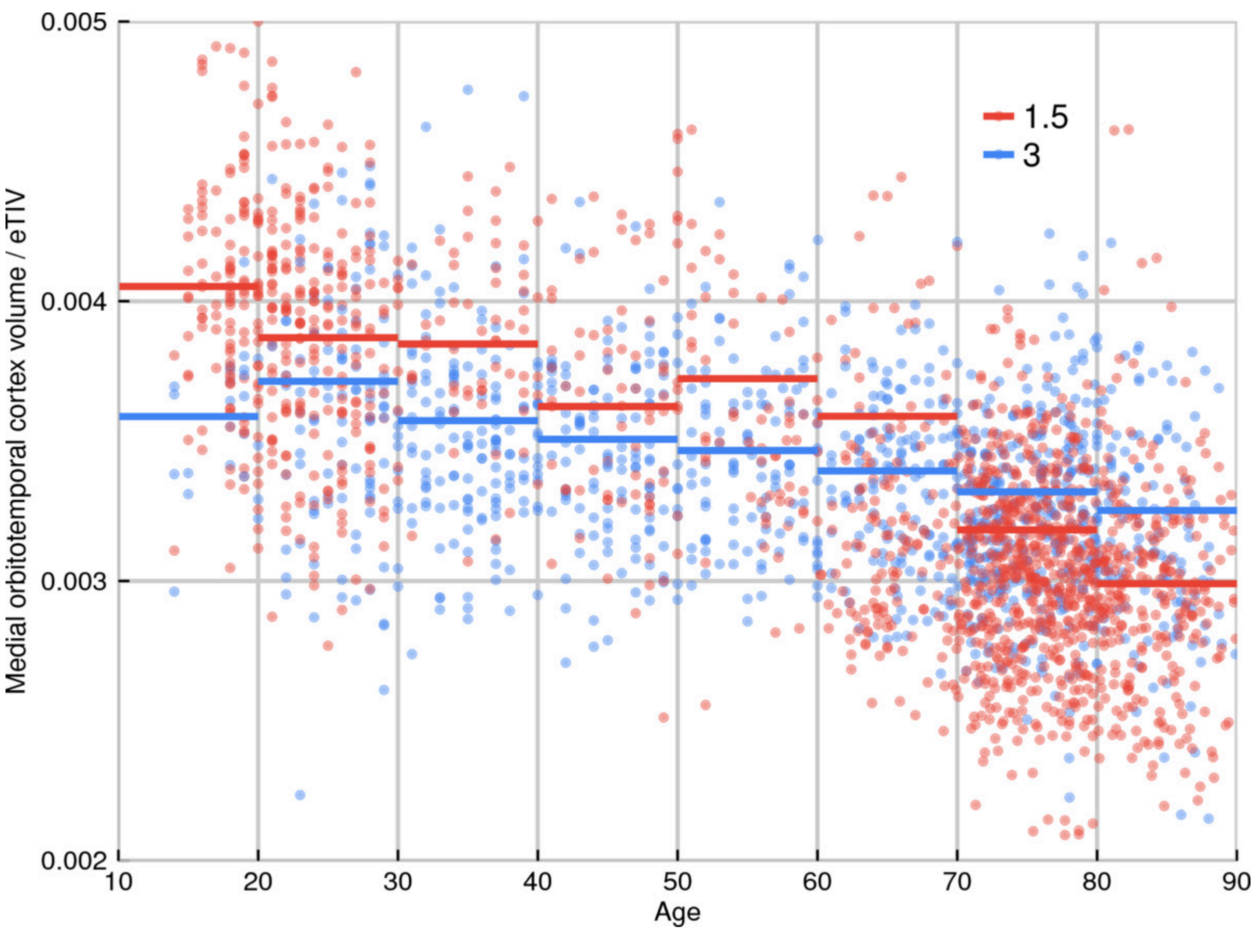

Fig. 3: Morphometric values change with age. Mean and standard deviation were obtained for each 10-year-bin

Results

The use of different acquisition protocols induces spurious patterns in the data. In other words, different protocols produce a bias in the morphologic measurements. Using random forest classifiers, we determined that the influence of the field strength is more significant than that of the MRI-machine manufacturer. Images obtained using different field strengths can be easily separated with this type of classifier. Figure 4 shows the 1.5 T and 3 T separation when applying a dimensionality reduction technique called t-SNE [5]. A harmonization algorithm was developed to minimize the influence of the field strength parameter: the z-score calculated for 1.5 T and 3 T for each 10-year-bin was the method that produced the best results in terms of compatibilization.

The effectiveness of the proposed solutions was measured in terms of the degree of confusion achieved during classification. Results showed that it is necessary to take into account brain measures variability with age. Best results were achieved when taking into consideration only people under 60 year old, reaching levels of confounding similar to the estimated chance accuracy.

Figure 5 shows the probabilities of class 3.0 T when using a random forest classifier over the test set, before (left) and after (right) the compatibilization method using data from people younger than 60 year old. For this group, accuracy dropped from 92% to 57%.

The same analysis was conducted with data from people older than 60 year old, as shown in Figure 6. In this case, accuracy did not decrease (below 89%).

The classification rate is substantially reduced after the compatibilization process for people under 60 years old but not for older people. However, rather than the accuracy of the classifier, it is the class probability for the test set that accounts for the effectiveness of the chosen approach. The probability distribution is narrowed and shifted towards 0.5 which can be thought as a proof for the classifier's increased uncertainty.

A feature importance analysis was conducted in order to determine which variables are not susceptible to the harmonization process. Indeed, we found that there is a group of features that are resistant to the compatibilization technique that was proposed. A preliminary analysis showed that some particular segmentation procedures from FreeSurfer are extremely dependent of the field strength: such as the estimation of blood vessels, white matter hypointensities. However these features are not closely linked to a cortical or subcortical structure but to the segmentation algorithms. Arguably, the differences in contrast produced by different field strengths affect FreeSurfer segmentation algorithms. Likewise, the PFE method demonstrated that some subcortical measurements (fissure of the hippocampus, thalamus, ventricles) as well as white matter estimations are important for the class distinction; meaning that their values are the most affected by the differences in the field strength.

The compatibilization process should not affect the power of classifying diseases that alter the brain structure. Confusion matrices in Figure 7 show the classification results for the data before compatibilization and for the harmonized data. In order to build the confusion matrix, we added up the test outputs for each fold. Apart from slightly reducing the number of false positives, applying the compatibilization methods does not affect the classification of pathologies. The feature importance analysis showed that the variables that were useful for classification were those with a well-known clinical relevance for AD, namely: hippocampal volumes as well as volumes and thickness from fusiform, inferior temporal and entorhinal cortex.

Images for this section:

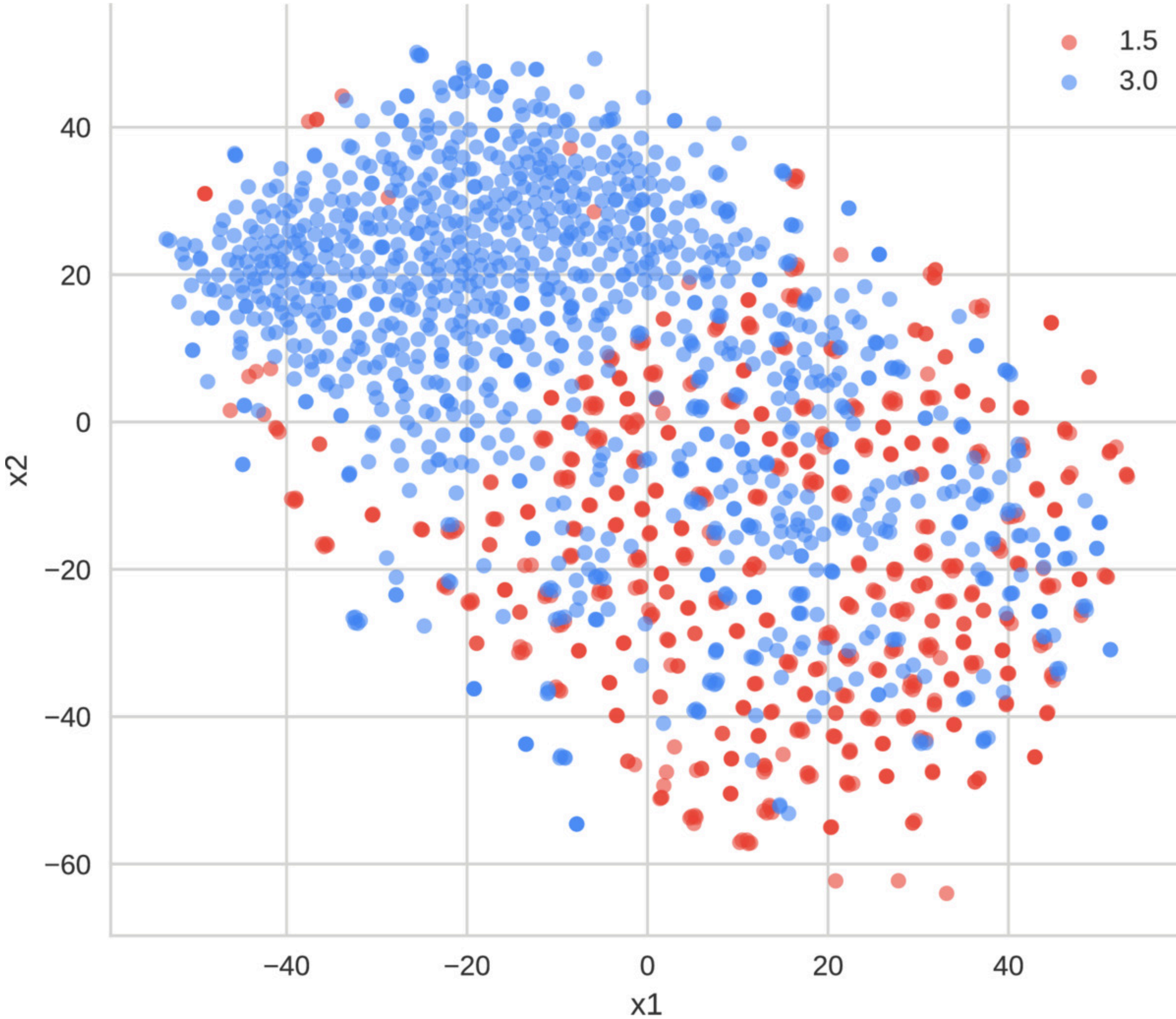

Fig. 4: t-SNE dimensionality reduction shows difference between groups

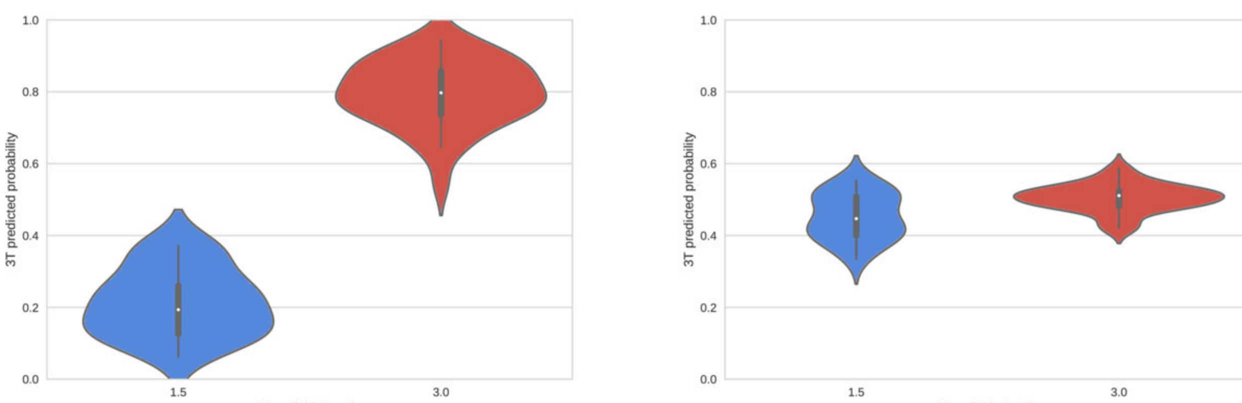

Fig. 5: Changes in the classification probability for people under 60

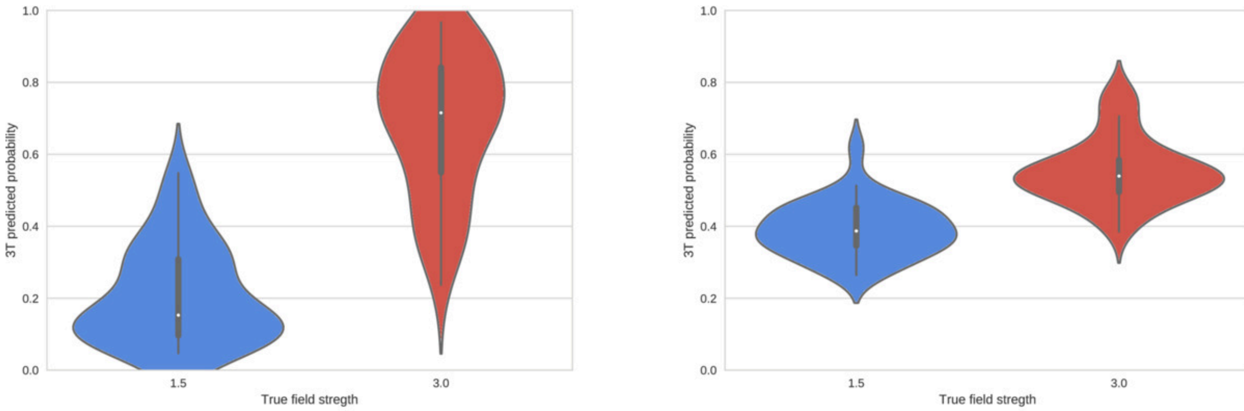

Fig. 6: Changes in the classification probability for people over 60

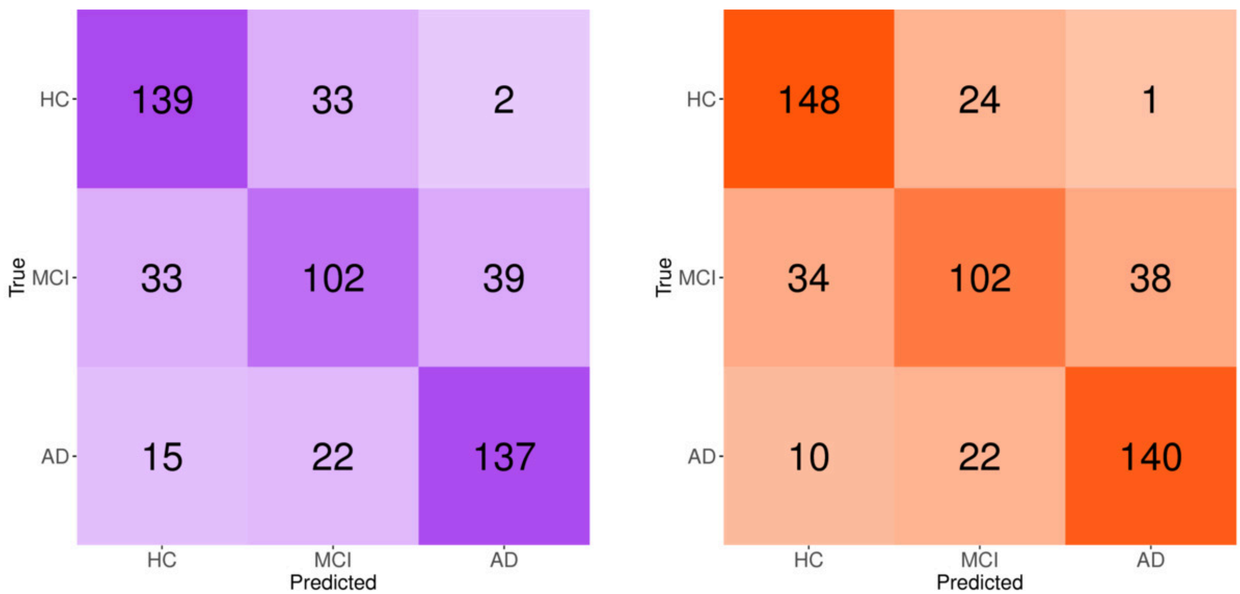

Fig. 7: Classification of diseases after (orange) the harmonization method is as effective as before (violet)

Conclusion

FreeSurfer's morphometric measurements are influenced by the MRI acquisition parameters. However, we introduced a method to remove those effects by using standardization techniques, allowing the translation of automatic quantitative MRI analysis from research to clinical applications.

We developed a technique for harmonizing data extracted from sMRI acquired using different field strengths and measured its effectiveness in terms of the ability to confound a random forest classifier. Our method consisted in calculating a 10-year-bin z-score for each class (1.5 T and 3.0 T). Our approach succeeds in reducing class separation without jeopardizing a classifier's ability to detect diseases.

Harmonization between field strength groups makes it easier to compare images from different sources, albeit there are a number of variables that cannot be used indistinctly. Since it is important to account for these features, an analysis of feature importance was conducted in order to determine which variables should not be analyzed indistinctly in spite of the harmonization process.

Finally our approach did not affect the classifier's capacity to differentiate AD, MCI and HC.

References

[1] Fischl, B. (2012). FreeSurfer. *Neuroimage, 62(2), 774-781.*

[2] Pedregosa, F. et al (2011). Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research, 12, 2825-2830.*

[3] P. A. Donnelly-Kehoe, G. Pascariello, J. C. Gomez. The changing brain in healthy aging: A multi-MRI machine and multicenter surface-based morphometry study. *In Proceedings of SPIE - The International Society for Optical Engineering. Volume 10160, 2017. Article number 101600B.*

[4] Donnelly-Kehoe, P. A., Pascariello, G. O., Gómez, J. C., & Alzheimers Disease Neuroimaging Initiative. (2018). Looking for Alzheimer's Disease morphometric signatures using machine learning techniques. *Journal of neuroscience methods, 302, 24-34.*

[5] Maaten, L. V. D., y Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(Nov), 2579-2605.