bioRxiv preprint doi: https://doi.org/10.1101/2020.10.23.352781; this version posted October 25, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MACHINE LEARNING FRAMEWORK FOR FULLY AUTOMATIC QUALITY CHECKING OF RIGID AND AFFINE REGISTRATIONS IN BIG DATA BRAIN MRI

Sudhakar Tummala[1]

[1]Department of Electronics and Communication Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

ABSTRACT

Rigid and affine registration to a common template is one of the vital steps during pre-processing of brain structural magnetic resonance imaging (MRI) data to make them suitable for further processing. Manual quality check (QC) of these registrations is tedious if the data contains several thousands of images. Therefore, I propose a machine learning (ML) framework for fully automatic QC of registrations by local computation of the similarity cost functions such as normalized cross-correlation, normalized mutual-information and correlation ratio, making them as features in training of ML classifiers. A MRI dataset consisting of 220 subjects from autism brain imaging data exchange is used for 5-fold cross-validation and testing. To facilitate supervised learning, the misaligned images were generated. Most of the classifiers reached testing F1-scores of 0.98 for checking rigid and affine registrations. Therefore, these ML models could be deployed for practical use.

***Index Terms*—** structural MRI, big data, supervised learning, quality control, image registration

1. INTRODUCTION

Due to advances in magnetic resonance imaging (MRI) data acquisition in bigger scale and making these data publicly open, researchers have been doing processing on larger number of subjects making them big data brain MRI studies [1, 2]. Some of these publicly available brain MRI big data sets include HCP (aging, young adult, development), IXI, ADNI, ABIDE, OASIS, UK biobank and NAKO Germany [3]. Due to advances in high performance computing and availability of these big data, brain studies on larger population is feasible making big data to big brain science [4, 5]. Structural MRI is the prime modality for the studies involving voxel based morphometry (VBM) of whole brain, surface based morphometry and profiling of deep gray matter structures [6-8].

Quality control (QC) has become an important step starting from assessing raw image quality till final post-processing. Checking the data for quality before pre-processing is the first stage of QC and there have been several studies proposing different metrics for quantitative assessment of raw image quality [9, 10].

After initial QC has passed, these acquired raw structural MRI images has to undergo several standard pre-processing steps such as reorientation to a standard template, cropping, bias correction and alignment of the images in subject space to a common space. Transforming images to a common space is one of the crucial steps in the pre-processing stage to make them better suitable for further processing such as VBM. These transformations could involve both rigid and affine registrations with 6 and 12 degrees of freedom respectively. QC of these registrations is necessary to make sure that they are suitable for subsequent analysis. In studies involving data in few hundreds, manual checking would be feasible and results in 100 percent accuracy. However, in big data studies that involve tens of thousands of images, manual inspection will be a tedious task and tiring as it is a time consuming process and moreover the manual process could be prone to inter/intra observer ratings especially with doubtful cases. Hence, this raises the need for fully automatic quality control mechanism which potentially reduce the manual intervention to the possible minimum.

Machine learning (ML) has been in wide use in medical imaging for several tasks such as assessment of image quality, brain mapping, and disease diagnosis and prognosis [11-13]. Coming to registration quality, the traditional similarity cost functions primarily in use are sum of squared differences, normalized cross-correlation (*NCC*), joint entropy based methods such as normalized mutual information (*NMI*) and correlation ratio (*CR*) [14]. However, computing them locally and combining them by employing ML classifiers may show better accuracy compared with accuracy using the individual cost functions alone.

To my knowledge, this is the first study aiming at development and comparison of different ML models for fully automatic QC of registrations. Hence, I propose to develop supervised ML models that could be deployed for use in big data structural MRI pre-processing.

2. METHODS

This section describes the dataset used for cross-validation and testing of the classifier models, generation of misaligned

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.23.352781; this version posted October 25, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

images for supervised learning, and computation of local cost values, and cross-validation and testing of different ML models.

2.1. Dataset

The data for cross-validation and testing consisted of a subset of high-resolution T1-weighted (MPRAGE) MRI images of 215 subjects extracted from publicly available autism brain imaging data exchange (ABIDE I). Out of 215 subjects, 110 subjects were with autism [2].

2.2. Pre-processing

The data was pre-processed under Nipype environment by calling the required third party interfaces such as ANTS, FSL, Freesurfer and other required software [15]. Firstly, the raw images were reoriented to a standard space and cropped to remove the neck region using the FSL tools *fslreorient2std*, *robustfov* respectively [16]. Further, the cropped images were bias corrected to remove the low-frequency intensity variations due to inhomogeneity of scanner magnetic field using the N4 method interface from ANTS [17]. Further, the images were aligned to the 1×1×1 mm^3 standard T1 template available in FSL both rigidly (that involves translation and rotation) and affinely (involves translation, rotation, scaling and shear) using the FSL *flirt* interface. After rigid and affine registration, the images were manually checked to make sure that they were correctly aligned to the template.

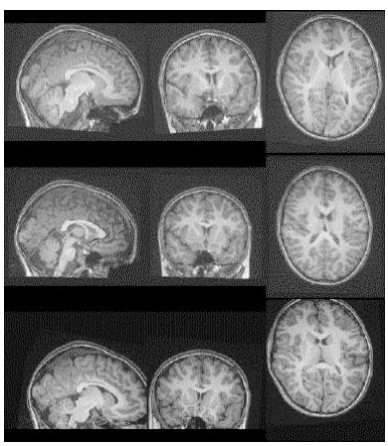

Fig. 1. Generated rigid misaligned images shown in slices in top (little variation) and bottom (more variation) rows, middle row is showing the correctly aligned image slices for comparison.

2.2. Generation of misaligned images

Since supervised learning approach was used in trained of ML models, misaligned rigid and affine test images were created by decomposing the corresponding transformation matrix into scales, translations and rotations; changing and applying these modified matrices to align the corresponding images in subject space to the common template. Ten test images (five each for rigid and affine) were generated for each image by different combinations of altering the matrix parameters to have variety of misaligned images. Fig. 1 showing the generated rigid misaligned images for a subject.

2.3. Computation of local cost values

Using custom written python script, the cost values that indicate the quality of registration between the transformed image and the template are computed locally using the cost functions such as *NCC*, *NMI* and *CR*. The computations were restricted to the brain region by using the brain mask of the corresponding 1×1×1 mm^3 standard FSL template. The computations were performed by defining a 3×3×3 volume of interest (VOI). The final cost value (*FCV*) is average of the cost values obtained by moving the VOI over the entire image with a stride of 3 and computing the cost values at each stride as described by the equation given below.

$$FCV = \frac{1}{C_n}\sum_{k=0}^{\frac{z-v_s}{s_t}}\sum_{j=0}^{\frac{y-v_s}{s_t}}\sum_{i=0}^{\frac{x-v_s}{s_t}} cost(M(a,\ b,c),T(a,b,c))$$

In above equation, $a = i * s_t{:}((i * s_t) + v_s)$, b $= j * s_t{:}((j * s_t) + v_s)$, and $c = k * s_t{:}((k * s_t) + v_s)$ indicate indices ranges of the moving image *M* and template *T* for defining the VOI. C_n is the total number of local cost computations, v_s is the size of VOI and s_t is the stride, and *cost* is the any one of the costs given in equations (1), (2), and (3) below. Lastly, *x*, *y*, and *z* represent the size of the template (or moving) in the three directions. Although, the size of VOI and stride are free parameters, they were restricted to 3. The cost *NCC* is computed as follows:

$$NCC = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i-\bar{x})(y_i-\bar{y})}{\sigma_x\sigma_y} \qquad \text{--- (1)}$$

Above, *N* is total number of voxels in either moving or template, x_i is the ith intensity and $\bar{x}$ is the mean intensity of the moving (image to be registered) and similarly, y_i is the is the ith intensity and $\bar{y}$ is the mean intensity of the template, σ_x and σ_y were standard deviations of intensities of moving image and template respectively. Further, the cost *NMI* is computed using the expression given below:

$$NMI(x;y) = \frac{I(x;y)}{\sqrt{H(x)H(y)}} \qquad \text{--- (2)}$$

Where $I(x;y) = H(x) + H(y) - H(xy)$ is mutual information between *x* (moving) and *y* (template), and $H(x)$ and $H(y)$ were entropy values of images *x* and *y* respectively. $H(xy)$ is the joint entropy of *x* and *y*. Finally, the cost *CR* is computed as follows:

$$\eta^2 = \frac{\sum_x n_x(\overline{y_x}-\bar{y})}{\sum_{x,i}(y_{xi}-\bar{y})} \qquad \text{--- (3)}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.23.352781; this version posted October 25, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In equation (3), y_{xi} is the ith intensity value in category x (category is either moving or template) n_x is the number of intensities in category x, $\overline{y_x}$ is average of category x and $\bar{y}$ is the mean of intensities of both moving and template. Finally, the *CR* is just η.

2.4. Training of ML classifiers

Python based Scikit-learn (*sklearn*) module was used for fitting/training of ML classifier models [18]. Before feeding to ML classifiers, the three features (cost values) *NCC*, *NMI* and *CR* were rescaled to have values between zero to one as a standard procedure. The ML classifiers that were cross-validated and tested include linear discriminant analysis (*LDA*), Gaussian naïve Bayes (*GNB*), linear support vector classifier (*SVC*), k-nearest neighbors (*kNN*, *k* randomly chosen as 15), random forest (*RF*, number of decision trees were randomly chosen as 100), and Adaptive Boosting (*AdaBoost* with 100 decision stumps).

The data was divided into two groups one for repeated (100 repetitions) stratified 5-fold cross-validation (*CV*) and the other group was used for testing (Table 1). Area under the ROC curve (AUC) and F1-score values were calculated using the *sklearn* module to indicate the quantitative performance of each classifier during *CV* as well as testing.

	Cross-validation N_1 (N_2)	**Testing N_1 (N_2)**
Correctly aligned	170 (87)	45 (23)
Misaligned	850 (435)	225 (115)

Table. 1. Number of images used for *CV* and testing. N_1 includes both healthy and diseased, N_2 is number of diseased.

3. RESULTS

The F1-scores during *CV* and testing for identifying the misaligned registrations using different ML classifiers is given in Table 2 for both rigid and affine.

Classifier	**F1-score (Rigid)**		**F1-score (Affine)**	
	CV	**testing**	***CV***	**testing**
LDA	0.947	0.982	0.956	0.981
SVC	0.917	0.914	0.935	0.922
GNB	0.916	0.967	0.915	0.977
kNN	0.971	0.988	0.960	0.982
RF	0.979	0.987	0.973	0.989
AdaBoost	0.977	0.984	0.971	0.986

Table. 2. F1-scores during *CV* and testing of different ML classifiers for checking both rigid and affine registrations.

The cross-validated ML models achieved the testing F1-scores in the range of 0.914–0.988 and the AUC values between 0.972–0.991 for identifying misaligned rigid registrations. Fig. 2 shows the true positive rate (sensitivity) and false positive rate (1-specificity) ROC curve for different cross-validated classifiers that were validated on the test set for finding faulty rigid registrations. Similarly, the cross-validated ML models reached the testing F1-scores and AUC values in the ranges of 0.922–0.989 and 0.989–0.993 respectively. Fig. 3 shows the ROC curve for different cross-validated classifiers that were validated on the test set for finding incorrectly aligned affine registrations.

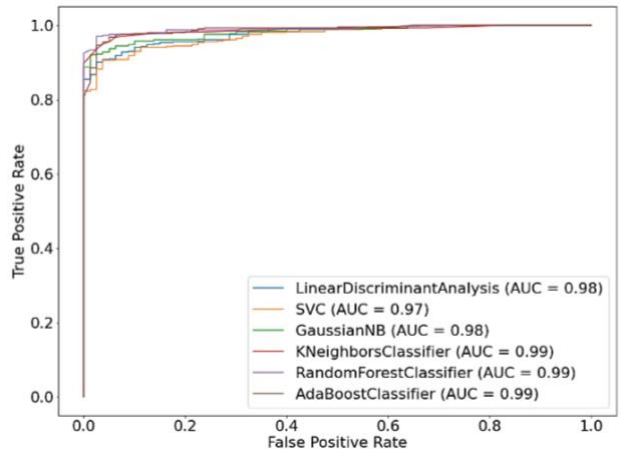

Fig. 2. Graph of true positive rate vs false positive rate for different ML classifiers, tested for rigid registrations.

4. DISCUSSION

Here, we developed a fully automatic framework using different ML classifier models for QC of rigid and affine registrations during structural MRI pre-processing. The classifier selection could be based on F1-score as the number of images in the two classes is heavily imbalanced. From Table 2, figs. 2 and 3, it is evident that all classifiers performed well, overall *kNN*, *RF* and *AdaBoost* classifiers demonstrated both better testing F1-scores and AUC values.

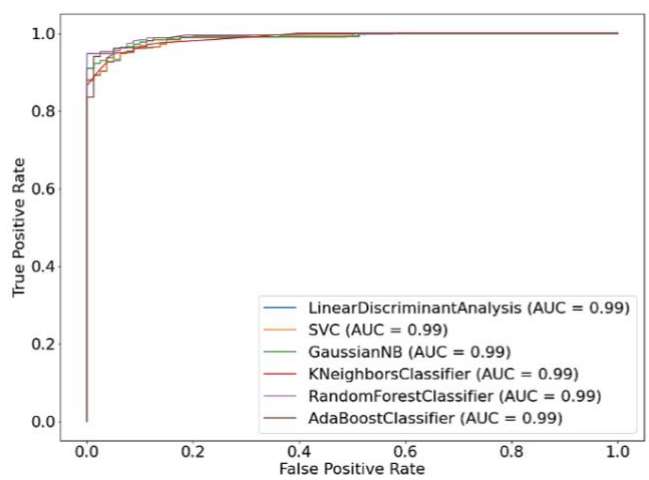

Fig. 3. Graph of true positive rate vs false positive rate for different ML classifiers, tested for affine registrations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.23.352781; this version posted October 25, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Another aspect to consider here is that checking the quality of alignment by computing the cost values locally using a moving VOI might have led to the anticipated F1-sores and AUC values on the test set. Also, since a stride of 3 is applied, which significantly reduced the computational time to around a minute for each local cost value, thus making it computationally efficient. All these cross-validated models could be saved using the python based *pickle* module for deployment.

5. CONCLUSION

The developed ML models could be deployed for fully automatic quality control of rigid and affine registrations in big data brain structural MRI pre-processing. Since the data is validated both on healthy and diseased brains, the trained models may be capable of identifying the misalignments both in health and disease. Future work could involve testing and improving these developed classifier models by using other cohorts and may be extended to deal with non-linear registrations. Further, the development of deep neural nets could be considered to eliminate the need for computation of cost values and the requirement of a template during training. Also, a single framework could be developed to deal with all types of structural images such as T1-weighted, T2-weighted, FLAIR and proton density.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by (ABIDE I). Ethical approval was not required as confirmed by the license attached with the open access data.

7. ACKNOWLEDGMENTS

The author(s) have no relevant financial or non-financial interests to disclose.

8. REFERENCES

[1] F. Alfaro-Almagro *et al.*, "Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank," *Neuroimage*, vol. 166, pp. 400-424, Feb 1 2018.

[2] A. Di Martino *et al.*, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Mol Psychiatry*, vol. 19, no. 6, pp. 659-67, Jun 2014.

[3] R. A. Poldrack and K. J. Gorgolewski, "Making big data open: data sharing in neuroimaging," *Nat Neurosci*, vol. 17, no. 11, pp. 1510-7, Nov 2014, doi: 10.1038/nn.3818.

[4] J. D. Van Horn and A. W. Toga, "Human neuroimaging as a "Big Data" science," *Brain Imaging Behav*, vol. 8, no. 2, pp. 323-31, Jun 2014.

[5] S. Chen *et al.*, "How Big Data and High-performance Computing Drive Brain Science," *Genomics Proteomics Bioinformatics*, vol. 17, no. 4, pp. 381-392, Aug 2019.

[6] R. Kumar *et al.*, "Brain putamen volume changes in newly-diagnosed patients with obstructive sleep apnea," *Neuroimage Clin*, vol. 4, pp. 383-91, 2014,.

[7] A. K. Bourisly, A. El-Beltagi, J. Cherian, G. Gejo, A. Al-Jazzaf, and M. Ismail, "A voxel-based morphometric magnetic resonance imaging study of the brain detects age-related gray matter volume changes in healthy subjects of 21-45 years old," *Neuroradiol J*, vol. 28, no. 5, pp. 450-9, Oct 2015.

[8] W. J. Henneman *et al.*, "Hippocampal atrophy rates in Alzheimer disease: added value over whole brain volume measures," *Neurology*, vol. 72, no. 11, pp. 999-1007, Mar 17 2009.

[9] L. Storelli *et al.*, "MRI quality control for the Italian Neuroimaging Network Initiative: moving towards big data in multiple sclerosis," *J Neurol*, vol. 266, no. 11, pp. 2848-2858, Nov 2019.

[10] O. Esteban *et al.*, "Crowdsourced MRI quality metrics and expert quality annotations for training of humans and machines," *Sci Data*, vol. 6, no. 1, p. 30, Apr 11 2019.

[11] M. L. Giger, "Machine Learning in Medical Imaging," *J Am Coll Radiol*, vol. 15, no. 3 Pt B, pp. 512-520, Mar 2018.

[12] M. N. Wernick, Y. Yang, J. G. Brankov, G. Yourganov, and S. C. Strother, "Machine Learning in Medical Imaging," *IEEE Signal Process Mag*, vol. 27, no. 4, pp. 25-38, Jul 2010.

[13] Y. Ding, S. Suffren, P. Bellec, and G. A. Lodygensky, "Supervised machine learning quality control for magnetic resonance artifacts in neonatal data sets," *Hum Brain Mapp*, vol. 40, no. 4, pp. 1290-1297, Mar 2019.

[14] W. R. Crum, T. Hartkens, and D. L. Hill, "Non-rigid image registration: theory and practice," *Br J Radiol*, vol. 77 Spec No 2, pp. S140-53, 2004.

[15] K. Gorgolewski *et al.*, "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python," *Front Neuroinform*, vol. 5, p. 13, 2011.

[16] S. M. Smith *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *Neuroimage*, vol. 23 Suppl 1, pp. S208-19, 2004.

[17] N. J. Tustison *et al.*, "N4ITK: improved N3 bias correction," *IEEE Trans Med Imaging*, vol. 29, no. 6, pp. 1310-20, Jun 2010.

[18] A. Abraham *et al.*, "Machine learning for neuroimaging with scikit-learn," *Front Neuroinform*, vol. 8, p. 14, 2014.