

FETMRQC : AUTOMATED QUALITY CONTROL FOR FETAL BRAIN MRI

Thomas Sanchez, Oscar Esteban ,Yvan Gomez , Elisenda Eixarch
, and Meritxell Bach Cuadra

PROBLEMATIC

- Fetal Brain MRI quality control is hard to automate :
 - Most automated QC have invalid prior (Adult brain)
 - Quality standard is different from adult brain MRI (ex : fetal motion)
- Fetal Brain Imaging often use on SSR
 - Construct 3D isotropic, high resolution, volumes with differently oriented 2D stacks
 - Rely on manual selection of stack

=> This study propose a QC tool on 2D low resolution stacks

DATA

- T2w stacks from 146 subjects
- 2 databases from 2 different institutions
 - CHUV : 61 subjects
 - BCNatal : 85 subjects
- 1010 LR series
- Two raters, 100 stacks in common.
 - Score between 1 and 4 for regression task
 - Pass/Fail label for classification

FEATURE EXTRACTION

Selected relevant IQM from MRIQC and implemented specific IQMs

Ex :

- mask volume, mask centroid
- Deep learning base score from other paper: dl_slice_iqa and dl_stack_iqa
- Propose new feature

Table 3. Detailed description of the metrics proposed for FetMRQC.

INTENSITY-BASED METRICS	
slice_loss	Use metrics commonly used for outlier rejection [13,14,18] to compute the difference between slices in the volume. We consider (normalized) mean averaged error, (normalized) mutual information, normalized cross correlation, (normalized) root mean squared error, peak signal-to-noise ration, structural similarity and joint entropy
sstats [8]	Compute the mean, median, standard deviation, percentiles 5% and 95%, coefficient of variation and kurtosis on brain ROI
entropy [8]	Measure the overall entropy of the image
bias	Level of bias estimated using N4 bias field correction [28]
filter_image	Estimate the sharpness by using Laplace and Sobel filters (commonly used for edge detection)
SHAPED-BASED METRICS	
closing_mask	Morphological closing of the brain mask in the through-plane direction, to detect inter-slice motion. Report the average difference with the original mask
filter_mask	Estimate the sharpness of the brain mask using Laplace and Sobel filtering. In an ideal case, the brain mask would be smoothly varying, especially in the through-plane direction

MODEL

- 2 Model for 2 tasks : classification and regression
- 3 variants :
 - Individual feature (best performing one is dl_stack_iqa)
 - Base : use a subset of feature (rank error, rank error full mask volume, centroid, centroid full, dl slice iqa and dl stack iqa)
 - FetMRQC : use every feature (including the one proposed)
- 2 Training setting :
 - in-domain for generalization on new subjects
 - out-of-domain for generalization on new sites

MODEL

Table 6. Selected hyperparameters for the different nested cross validation procedures. The in-domain experiment uses 5-fold nested cross-validation, while the out-of-domain experiment splits data by site (CHUV and BCNatal) and as a result has only two folds. The list of possible parameters is provided in Table 5.

In-domain – Regression						In-domain – Classification					
	Remove feat.	Scaling	Winnow	PCA	Model		Remove feat.	Scaling	Winnow	PCA	Model
BASE	✗	Standard	✓	✓	Random Forest	BASE	0.8	Robust	✓	✓	Random Forest
	0.9	Standard	✓	✗	Random Forest		✗	None	✓	✓	Random Forest
	0.8	Standard	✓	✗	Random Forest		✗	Standard	✓	✗	Random Forest
	0.8	Standard	✓	✗	Gradient Boosting		0.8	Standard	✓	✗	Random Forest
	0.8	Standard	✓	✗	Random Forest		✗	Standard	✓	✗	Random Forest
FETMRQC	✗	Robust	✓	✓	Random Forest	FETMRQC	✗	Standard	✓	✓	Gradient Boosting
	✗	Robust	✓	✓	Gradient Boosting		0.9	Robust	✓	✓	Random Forest
	0.8	None	✓	✗	Random Forest		0.8	Robust	✓	✗	Gradient Boosting
	0.9	Robust	✓	✗	Gradient Boosting		0.9	Standard	✓	✓	Lin. Regression
	✗	None	✓	✗	Random Forest		✗	Standard	✓	✗	Gradient Boosting
Out-of-domain – Regression						Out-of-domain – Classification					
	Remove feat.	Scaling	Winnow	PCA	Model		Remove feat.	Scaling	Winnow	PCA	Model
BASE	0.9	Robust	✓	✗	Gradient Boosting	BASE	0.8	Standard	✓	✗	Random Forest
	✗	None	✓	✓	Random Forest		0.8	Standard	✓	✓	Random Forest
FETMRQC	0.9	None	✓	✓	Lin. Regression	FETMRQC	0.9	Standard	✓	✗	Random Forest
	✗	Robust	✓	✗	Random Forest		0.8	Standard	✓	✗	Random Forest

Table 1. In-domain evaluation

	Regression		Classification	
	MAE (\downarrow)	Spearman (\uparrow)	F1 (\uparrow)	AUC (\uparrow)
dl_stack_iqa	0.72 ± 0.05	0.37 ± 0.11	0.85 ± 0.02	0.53 ± 0.02
Base	0.58 ± 0.05	0.60 ± 0.03	0.88 ± 0.02	0.71 ± 0.06
FetMRQC	0.53 ± 0.09	0.71 ± 0.05	0.90 ± 0.02	0.77 ± 0.07

Table 2. Out of domain evaluation

	Regression		Classification	
	MAE (\downarrow)	Spearman (\uparrow)	F1 (\uparrow)	AUC (\uparrow)
dl_stack_iqa	0.75 ± 0.03	0.42 ± 0.10	0.83 ± 0.04	0.57 ± 0.07
Base	0.75 ± 0.01	0.38 ± 0.03	0.85 ± 0.02	0.67 ± 0.06
FetMRQC	0.68 ± 0.06	0.50 ± 0.08	0.89 ± 0.01	0.77 ± 0.05

RESULTS

IMPACT ON SRR

- NiftyMIC and NeSVoR are two super resolution reconstruction techniques

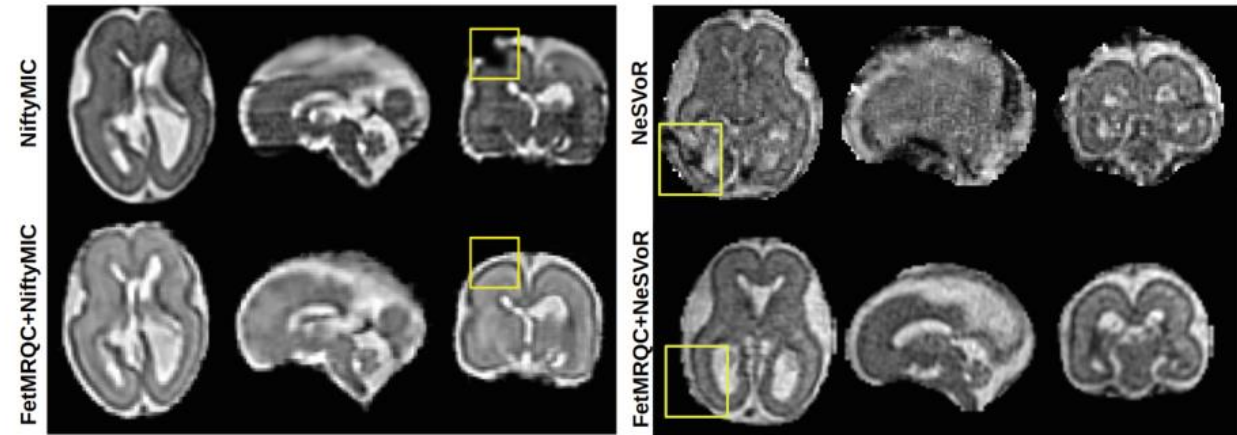


Fig. 7. Illustration of usefulness of QC on the SR reconstruction, using NiftyMIC [18], and NeSVoR [17]. FetMRQC+NiftyMIC successfully removes 6 series out of the 13 available, and FetMRQC+NeSVoR removes 2 out of 5. FetMRQC leads to a substantially greater image quality.

CONCLUSION

- The paper claim that :
 - QA and QC for T2w fetal brain MRI can be automated
 - Classical machine learning help to have interpretable decision and good generalization
- Issue :
 - No specific “winning” architecture mentioned
 - Deep Learning based feature seems to really important, even more for classification, in which measure are they interpretable.