

Protocol for the quality rating of 3D super-resolution reconstruction in fetal brain MRI

Meritxell Bach Cuadra, Thomas Sanchez, Gerard Marti Juan,
Angeline Mihailov, Guillaume Auzias



Lausanne, 20/08/25

CIBM.CH

Research context

These guidelines are written in the context of the [MULTIFACT ERANET NEURON project](#). Our multi-centric project aims at developing machine learning methods for the study of large-scale fetal brain Magnetic Resonance Images (MRI), with the goal to provide researchers with advanced tools and image analysis methods to study early brain development. Specifically, we develop tools from super-resolution reconstruction algorithms to tissue segmentation, surface extraction, and normative modelling. Our ultimate goal is to be able to depict abnormal trajectories with the extracted biomarkers in the context, for instance, of corpus callosum agenesis or intra-uterine growth restriction.

Quality control (QC) of input images is a crucial step in any image analysis pipeline. It is very well known that **bad quality data can strongly bias the extracted imaging biomarkers**. In the context of large-scale fetal brain MRI studies, QC is key to be performed before [1] and after the super-resolution reconstruction (SRR) step.

Quantifying SRR image quality is key as it has been recently demonstrated that is one of the major domain shifts factors influencing semantic segmentation [2].

However, existing MR quality tools for adults [3], [4] or designed for original low-resolution T2-weighted fetal brain sequences [1], [5], [6] cannot be directly applied to 3D SRR fetal brain MRI.

Indeed, the SRR processing step may introduce specific artefacts such as topological errors or regularization artifacts (see Figure 1C) that deserve specific assessment [7].

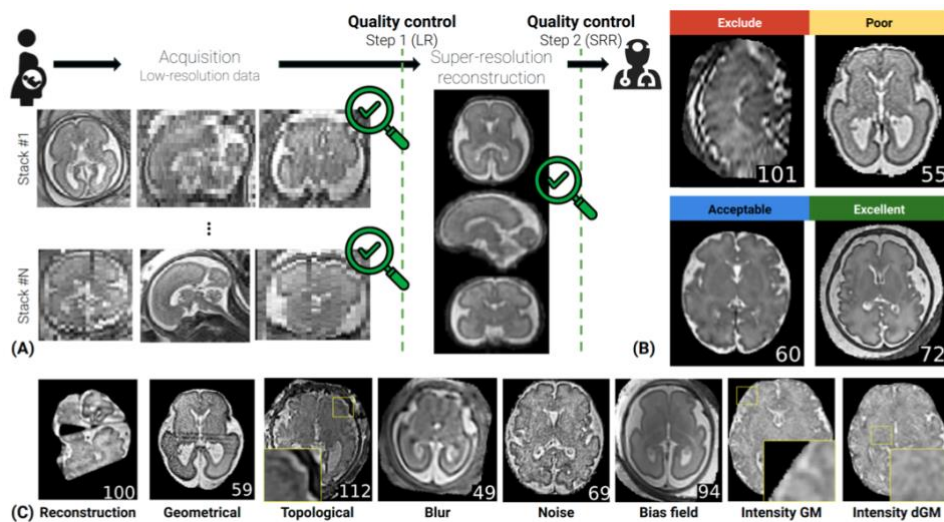


Figure 1. Illustration of SRR pipeline (A) and different SRR output quality (B) and specific artefacts (C).

If you use this report

The proposed protocol is derived from previous works on fetal and adult brain MRI quality control [1], [3] and is a companion to our work “Assessing Data Quality on Fetal Brain MRI Reconstruction: A Multi-site and Multi-rater Study,” published at PIPPI 2024 doi: [10.1007/978-3-031-73260-7_5](https://doi.org/10.1007/978-3-031-73260-7_5) [7].

If you use this protocol, please cite:

Bach Cuadra, M., Sanchez, T., Martí Juan, G., Mihailov, A., Auzias, G. (2025) “Protocol for the quality rating of 3D super-resolution reconstruction in fetal brain MRI.” Zenodo. <https://doi.org/10.5281/zenodo.15696638>

This protocol was used already used in the following works:

- Zalevskyi, V., Sanchez, T., Kaandorp, M., Roulet, M., Fajardo-Rojas, D., Li, L., ... & Cuadra, M. B. (2025). Advances in Automated Fetal Brain MRI Segmentation and Biometry: Insights from the FeTA 2024 Challenge. arXiv preprint arXiv:2505.02784.
- Sanchez, T., Zalevskyi, V., Mihailov, A., Martí-Juan, G., Eixarch, E., Jakab, A., ... & Cuadra, M. B. (2025). Automatic quality control in multi-centric fetal brain MRI super-resolution reconstruction. arXiv preprint arXiv:2503.10156.

General remarks

We aim at **standardizing the QC manual annotation** procedure (as to reduce inter-rater variability) for the evaluation of QC **in 3D SRR fetal brain MRI**.

Ideally, we would recommend performing the annotations independently of the many possible downstream tasks like fetal brain biometry measurements or semantic segmentation for instance.

Focus should be on the brain area, despite some SRR methods may have a halo or even reconstruct mother tissues.

We need to explore images as 3D volumes, observing quality as visible in the three different Axial, Sagittal and Coronal views.

Based on our previous works, we recommend **rating specific image artefacts first** (e.g. blurriness, contrast, topology etc.) and **at the end only provide one global rating**.

Ensure you use a screen with at least 24" and good light conditions in the room.

Generating the reports

The generation of the reports is described in detail at https://github.com/Medical-Image-Analysis-Laboratory/fetmrgc_sr. It requires installing a python library and generating html reports following the listed instructions.

Annotating reports

Manual annotations are executed through a browser using html widget adopted from [1], [3] and illustrated in Figure 2. Some further comments:

- Use Chrome preferably (Safari has some bugs)
- Load index.html
- Scroll all thumbnail images, note that sagittal and coronal views are at the bottom
- Brain orientation has to be in the regular radiological format.
- You can then turn on the rating widget (top right)
- You can proceed with your evaluation, first rate if the fetal brain is fully reconstructed, that is, no brain area has been cut.
- When opening the widget, if you don't visualize it properly or it is too big or too small, you can change the zoom of the webpage. The images will remain the same size.
- Download report (ensure you know where they are saved)
- If you plan on doing the evaluation not all at once, it's better to save the progress by clicking on "Save progress config" at "Home" page when you finish the subset -- This will download a *.json* file that you can later upload to the "Home" page by clicking on "Choose File" button to recover the progress (those in green are done)
- If you forgot to save the progress, you could always check your last downloaded *.json* file name

[Choose File](#) [Save progress config](#)

Index of reports

Rated cases: 0/180

- X sub-YXEHX_report.html
- X sub-HZMCT_report.html
- X sub-CPZSS_report.html
- X sub-FIAUY_report.html
- X sub-BTMTQ_report.html
- X sub-EQHAB_report.html
- X sub-IZUFM_report.html
- X sub-RCRPU_report.html
- X sub-JJILJ_report.html
- X sub-DCXPU_report.html
- X sub-JPLJE_report.html
- X sub-DUXFP_report.html
- X sub-VZJFO_report.html

Rate Image

Is the brain fully reconstructed?

Geometrical artefacts (e.g. stripes)

Topological artefacts (e.g. discontinuous cortical ribbon, cortical CSF in contact with WM, etc.)

Noise

Tissue contrast - I (intensity at WM/cortical GM/CSF interface)

Tissue contrast - II (intensity at WM/Deep GM interface)

Global subjective quality rating:

Extra details

[Download](#)

[Previous](#) [Home](#) [Next](#)

☒ Rating widget

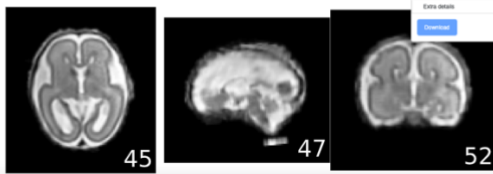
Super-resolution reconstruction report

Summary

- BIDS filename: sub-YXEHX
- Image resolution: 256 x 256 x 256
- Voxel size: 1.125 x 1.125 x 1.125 mm³

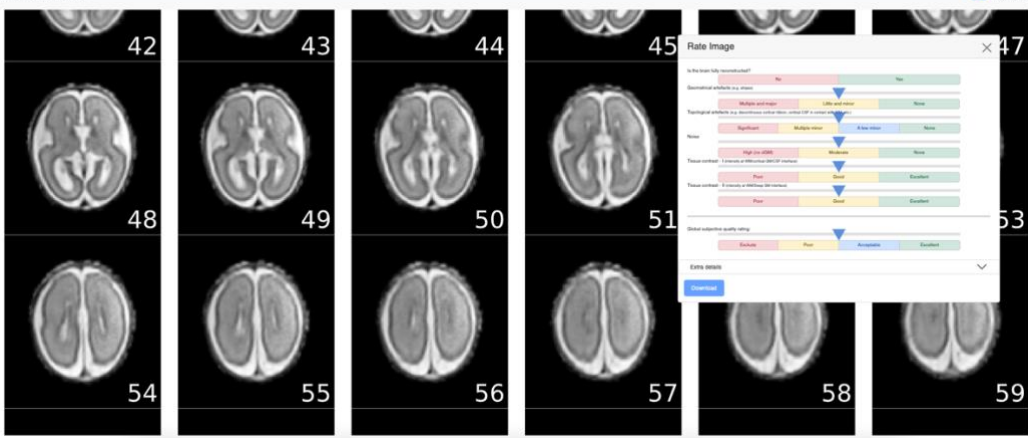
Visual reports

Quick view



[Previous](#) [Home](#) [Next](#)

☒ Rating widget



Rate Image

Is the brain fully reconstructed?

Geometrical artefacts (e.g. stripes)

Topological artefacts (e.g. discontinuous cortical ribbon, cortical CSF in contact with WM, etc.)

Noise

Tissue contrast - I (intensity at WM/cortical GM/CSF interface)

Tissue contrast - II (intensity at WM/Deep GM interface)

Global subjective quality rating:

Extra details

[Download](#)

Figure 2. Different windows of visual reports HTML. Top left index, top right zoom on widget for evaluation, middle and bottom row view of the thumbnails.

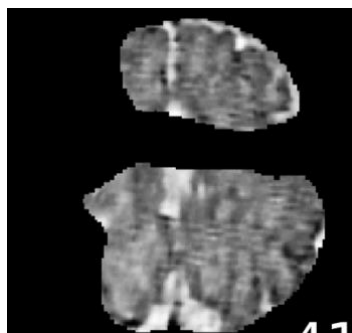
Description of specific image artefacts

We do only include here the specific image artefacts we found consistent within raters and across raters to be annotated in our study at PIPPI 2024 [7].

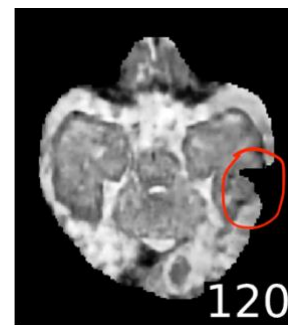
Is the brain fully reconstructed?

We might observe that some brain areas have been cut significantly, indicate yes to the widget and this will lead also to the grading of topological artefacts, it would certainly imply exclusion of further analysis. For instance, the example image of the left here below obviously failed the SRR step and was not fully reconstructed. More subtle cuts may though appear still are relevant to be indicated and may be reason for exclusion in some scenarios.

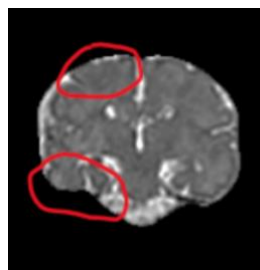
Obvious bad quality



Subtle bad quality: some small parts of the brain are missing due to inaccurate masking

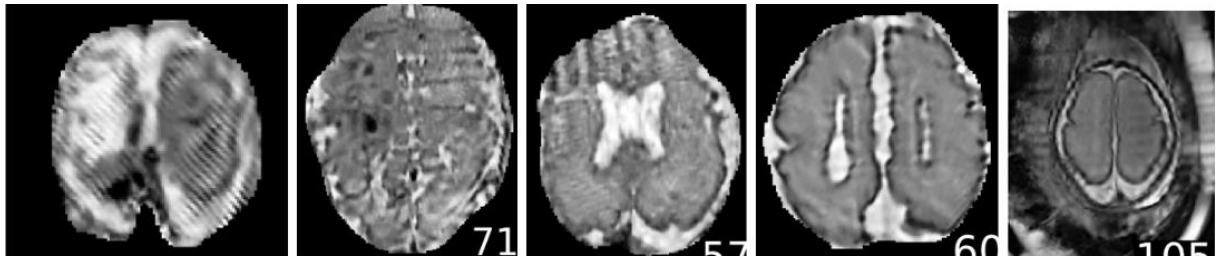


We do not consider subtle, ambiguous boundaries as warranting a rating of “not reconstructed” on the brain. However, a case like the one below might be assigned a “reconstructed” score but be penalized at the level of topological consistency.



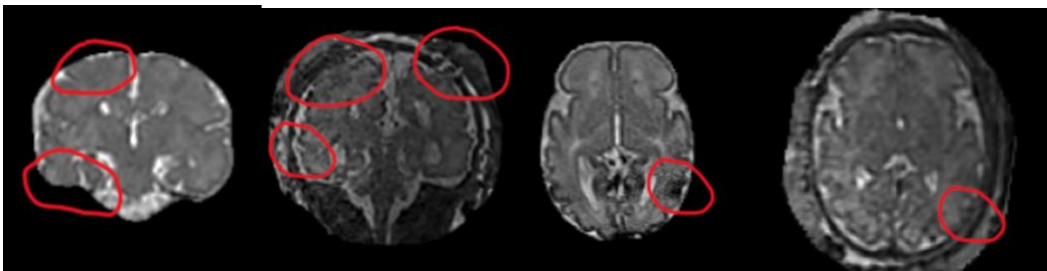
Geometrical artefacts

These artefacts refer to repetitive patterns such as lines or checkboard or multiple dark holes that are rather related to regularization “issues” of the SRR methods. First 3 examples below have strong geometric artefacts (multiple and major), the other 2 most right clearly present artifacts as well but not so strong (minor).



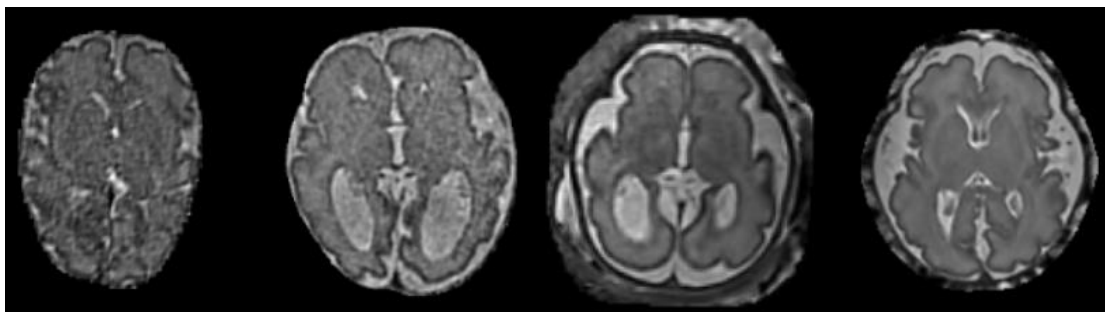
Topological artefacts

Topological artefacts refer to cases where image presents non plausible anatomical configurations. For instance, WM directly over the background, disconnected/broken GM, etc. Here below we illustrate several examples with topological artefacts.



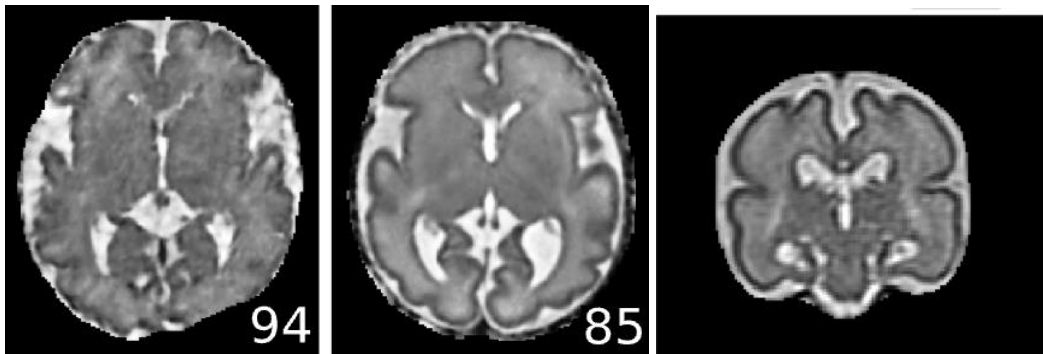
Noise

The presence of noise in the image locally or uniformly distributed may appear like *salt and pepper* noise in the SRR volume. Two left images have high level of noise, while too on the right do not have. We often focus on the central GM area of thalamus and caudate/putamen to know if noise is hampering the direct visualization of central nuclei structures.



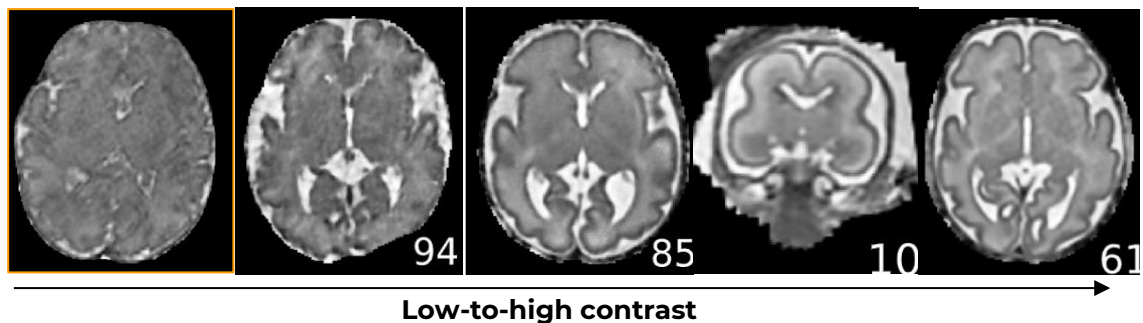
Intensity contrast I – Look at cortical interfaces WM/GM and GM/CSF

Focus to be paid at outer anatomy to identify if the contrast between CSF, GM and WM is pronounced, often we aim at having a very dark/black cortical ribbon. Low contrast may be due to significant blurry aspect and /or topological errors of the cortical ribbon reconstruction (left image here below) or due to a more subtle contrast difference (middle panel). The right panel presents a very good cortical GM contrast.



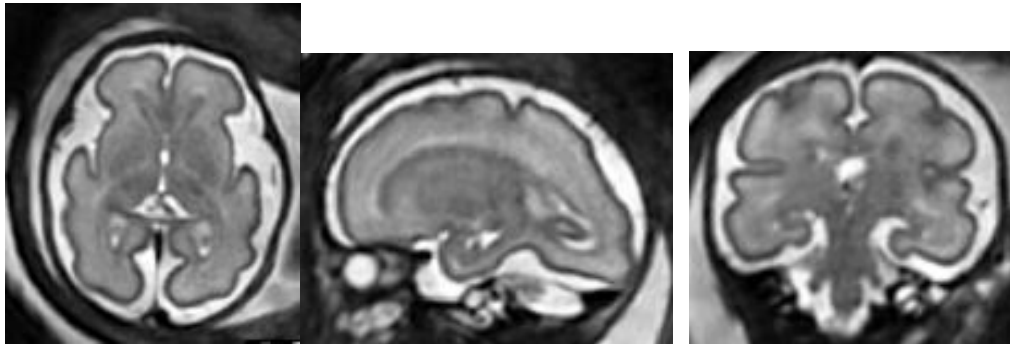
Intensity contrast II – Look at deep structures interfaces WM/GM

We here aim at evaluating rather the overall WM contrast variations towards its center and in particular if we can well distinguish the border of the central GM structures such as germinal matrix (early GA), thalamus and internal capsule (at later GA stages). From left to right and top to bottom: low deep GM/WM contrast to high.



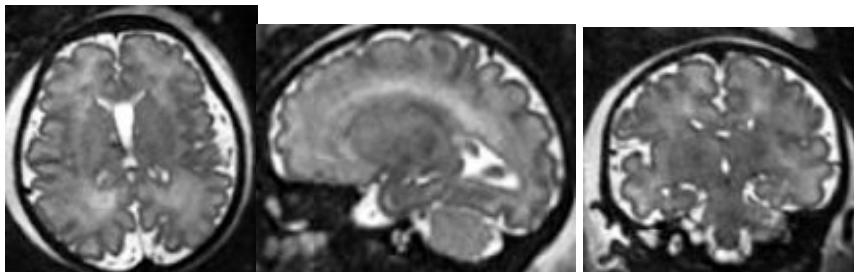
Examples of global quality ratings

Excellent quality



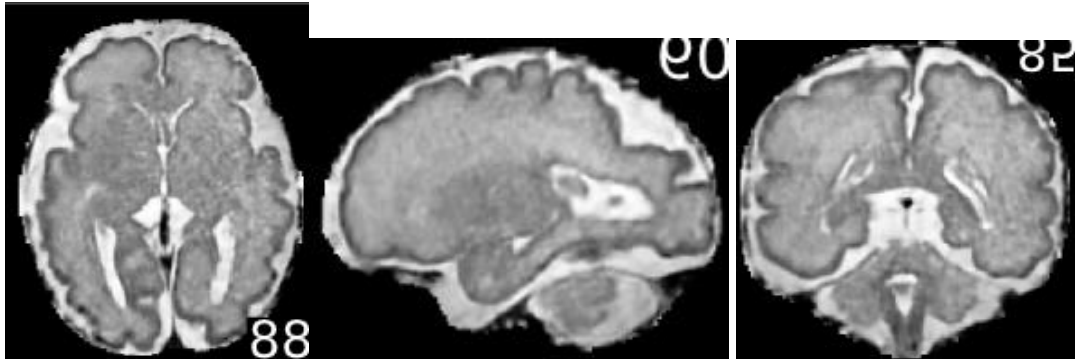
This brain is considered top excellent quality for several reasons including that it is completely reconstructed and there are no major topological artefacts anywhere along the cortex or ventricles. Furthermore, there is no noise which allows for easy distinction between the gray matter and CSF, between the gray and white matter and within the white matter (heterogeneous maturation and myelination). There is also very minimal blurring and nearly no bias.

Acceptable to Excellent



The brain above is within the “acceptable-to-excellent” range. This brain is fully reconstructed, with gray matter-CSF and gray matter-white matter contrasts clearly delineated. Furthermore, the white matter components (such as subcortical regions) are well illustrated. The reason this brain is not considered fully “excellent” is because, although fully visible, the border contrasts are slightly darker and not as clear as in the “excellent” case above. There is also a minimal amount of noise and a few topological errors (particularly around the orbitofrontal area, as seen in the sagittal axis).

Acceptable



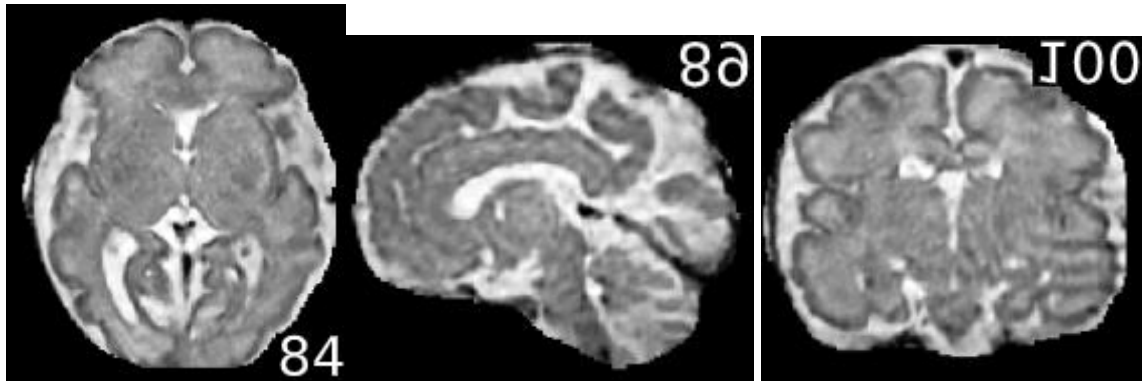
The brain above is considered acceptable (not excellent) due to many reasons including the fact that despite being fully reconstructed, there is obvious noise. This noise affects the clear delineation of the gray matter-CSF and gray matter-white matter borders. It further affects the distinction of the white matter subregions. There are also no major topological artefacts.

Bad quality (Middle poor)



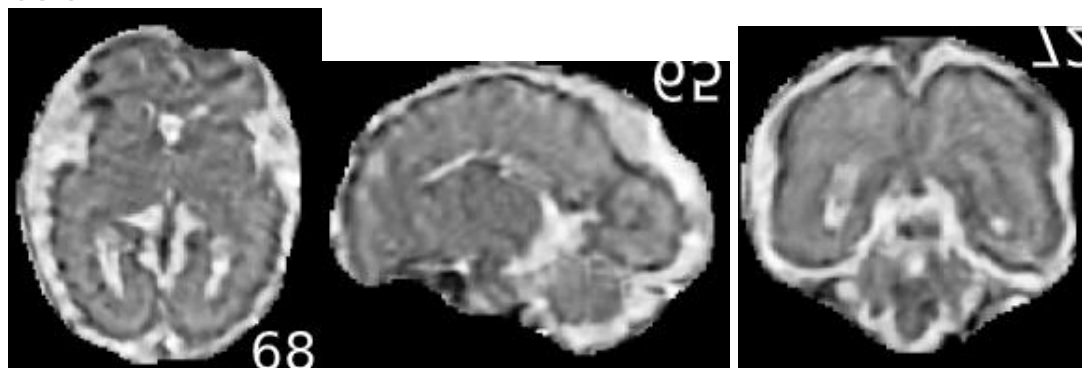
This scan is labeled as “middle-poor” since though the brain is fully reconstructed with no geometrical artefacts, there are a few topological artefacts as well as a lot of noise, which diminishes greatly the gray matter-CSF and gray matter-white matter contrasts. The noise also significantly limits the delineation of subcortical structures and white matter development and/or myelination.

Bad quality (poor at the border of exclusion)



This scan is labeled as “poor-to-exclude” since there are obvious topological artefacts (cf. sagittal plane) resulting in a small piece of missing brain as well as obvious geometric artifacts (cf. coronal plane). There is also noise which limits the sharpness of subcortical white matter structures as well as CSF-gray matter and gray matter-white matter contrasts.

Exclusion

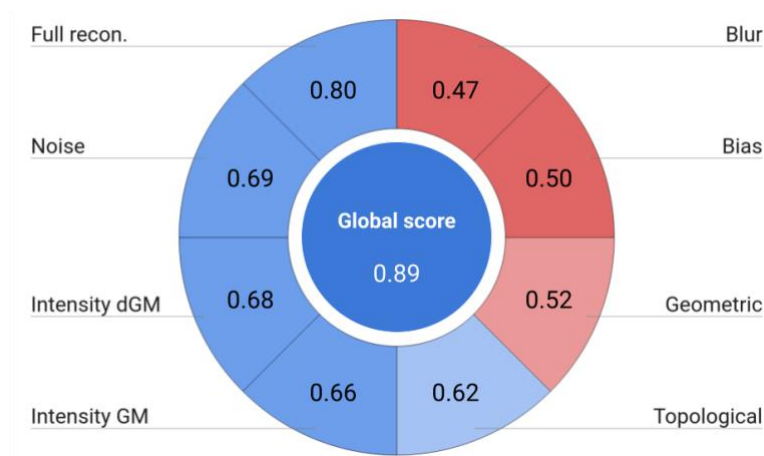


This scan would be excluded since there is an obvious topological artefact in the form of missing brain. Furthermore, there is a lot of noise and blurred parts of the brain, resulting in a poor gray matter-CSF and gray matter-white matter contrast. Subcortical and other white matter features are almost absent. Though subtle, there are also geometrical artefacts.

Discussion

In our experiments, this protocol allowed us to achieve a high intra- and inter-rater reliability on the overall quality score, around 0.85-0.90 in reliability.

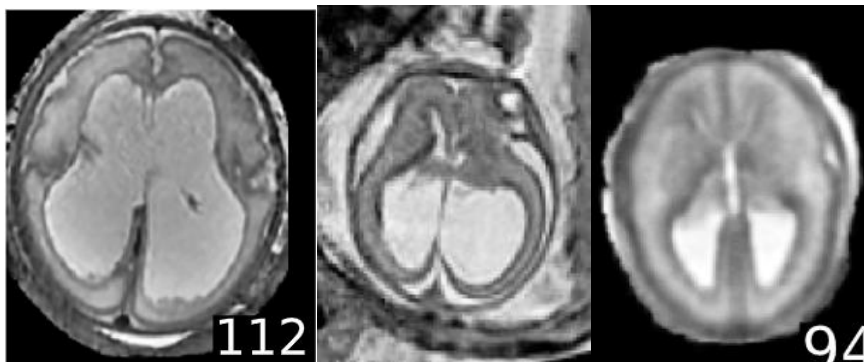
While rating specific artifacts leads to more intra- and inter-rater variability, reading the guidelines and agreeing among raters by reviewing a few cases helped improve the agreement. Nevertheless, to simplify rating, we did not include in this protocol the rating of blur and bias field, as we found both metrics to be rated unreliably, as illustrated below.



Even if you are interested in rating only the global score, we generally observed that following the individual specific artefact ratings first was important to reach high agreement in global evaluations.

Some difficulties remain nevertheless for future works.

- Some artefacts might be biased due to gestational age, for instance cortical ribbon may appear darker and less blurred in younger fetuses, as T2 values and cortical folding undergo significant changes during gestation.
- Pathological cases that present a huge anatomical variability where some areas might be difficult to see (deep GM due to large ventriculomegaly, spina bifida no cortex visible) will influence the ratings.



Acknowledgements

This work was funded by the Era-Net Neuron MULTIFACT project, under local project numbers SNSF 31NE30_203977 (TS), ANR-21-NEU2-0005 (GA, AM), PCI2021-122044-2A (GMJ). GA, AM acknowledge support from the SulcalGRIDS project (ANR-19-CE45-0014). TS and MBC acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by CHUV, UNIL, EPFL, UNIGE and HUG.

References

- [1] "FetMRQC: A robust quality control system for multi-centric fetal brain MRI," *Med. Image Anal.*, vol. 97, p. 103282, Oct. 2024, doi: 10.1016/j.media.2024.103282.
- [2] V. Zalevskyi *et al.*, "Advances in Automated Fetal Brain MRI Segmentation and Biometry: Insights from the FeTA 2024 Challenge," arXiv.org. Accessed: May 27, 2025. [Online]. Available: <https://arxiv.org/abs/2505.02784v3>
- [3] O. Esteban, D. Birman, M. Schaer, O. O. Koyejo, R. A. Poldrack, and K. J. Gorgolewski, "MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites," *PLOS ONE*, vol. 12, no. 9, p. e0184661, Sept. 2017, doi: 10.1371/journal.pone.0184661.
- [4] M. Garcia, N. Dosenbach, and C. Kelly, "BrainQCNet: a Deep Learning attention-based model for multi-scale detection of artifacts in brain structural MRI scans," Mar. 14, 2022, *bioRxiv*. doi: 10.1101/2022.03.11.483983.
- [5] J. Xu *et al.*, "Semi-Supervised Learning for Fetal Brain MRI Quality Assessment with ROI consistency," *Med. Image Comput. Comput. Assist. Interv. – MICCAI 2020*, June 2020, Accessed: May 10, 2022. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-59725-2_37
- [6] L. Liao *et al.*, "Joint Image Quality Assessment and Brain Extraction of Fetal MRI Using Deep Learning," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, vol. 12266, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racocceanu, and L. Joskowicz, Eds., in Lecture Notes in Computer Science, vol. 12266, Cham: Springer International Publishing, 2020, pp. 415–424. doi: 10.1007/978-3-030-59725-2_40.
- [7] T. Sanchez *et al.*, "Assessing Data Quality on Fetal Brain MRI Reconstruction: A Multi-site and Multi-rater Study," in *Perinatal, Preterm and Paediatric Image Analysis*, D. Link-Sourani, E. Abaci Turk, C. Macgowan, J. Hutter, A. Melbourne, and R. Licandro, Eds., Cham: Springer Nature Switzerland, 2025, pp. 46–56. doi: 10.1007/978-3-031-73260-7_5.