SIIM CAIMI26 Conference on Artificial Intelligence in Medical Imaging
OCT 26-27 | University of Pennsylvania, Philadelphia, PA

NeuroQC-Assist: AI-Assisted Neuroimaging Quality Control

Sai Krishna C. Annavazala, MS, University of Pennsylvania; Nima Broomand Lomer, MD; Tyler Spears, PhD; Drew Parker; Ragini Verma, PhD

Briefly describe what attendees will see during your demo.
During the demo, attendees will see the NeuroQC-Assist workflow using sample T1-weighted MRI scans. The demo will begin with image preparation, showing how each scan is conformed to a fixed input size by locating the brain and cropping, and how a subject-space brain mask is generated. The prepared image and brain mask are then passed into the image-quality classification model. For each scan, the workflow produces a poor-quality confidence score ranging from 0 to 1, indicating the confidence in grading the image as good (0–0.5) or bad (0.5–1), and a structured CSV file with the results. The demo will also show how confidence scores are used in a human review workflow to classify scans as PASS, REVIEW, or FAIL. Finally, Grad-CAM overlays will be shown next to the original MRI views. These overlays provide a rough visual indication of which image regions influenced the image-quality classification model's prediction, especially for scans marked REVIEW or FAIL.

Problem & Motivation: What clinical or operational problem does your MVP address?
Neuroimaging scans often need to be checked before they can be used for analysis, but this review is still commonly done manually. That can be manageable for a small number of scans, but it becomes slow and inconsistent when working with large datasets, multiple sites, and different acquisition settings. Poor-quality scans can also create problems later in the analysis. They may lead to failed preprocessing, inaccurate brain masks or segmentations, unreliable measurements, or unusable downstream model outputs. These issues are especially frustrating when they are discovered only after substantial time has already been spent processing the scan. Manual quality control is also typically performed after the scanning session, when the opportunity to reacquire a better image may already be lost. Performing automated quality control as close to the time of acquisition as possible could help identify problems early enough to repeat the scan and preserve valuable scanner sessions. The motivation for this MVP is to make the first quality-control pass easier, more consistent and scalable. The goal is not to replace expert review, but to help prioritize which scans can move forward, which scans are likely poor quality, and which scans need a closer look.

What You Built: Describe your MVP, tool, or workflow in plain language
We built a containerized workflow for quality control of T1-weighted MRI scans. The current version is a working prototype that takes an input scan, prepares it in a consistent format, runs an image-quality model, and saves the results in a spreadsheet that can be reviewed. The workflow begins with a raw T1-weighted MRI scan. It conforms the scan into a fixed input size and generates corresponding brain mask. The model is a three-dimensional convolutional neural network designed to evaluate scan quality. For each scan, it produces a confidence score for poor image quality and a good-or-bad prediction. NeuroQC-Assist then organizes the result into a practical quality-control category. The main output is spreadsheet file the scan name, model score, prediction, and quality-control category. The workflow can also save Grad-CAM overlays

for selected cases, which can be viewed next to the original MRI images during review.

Tech Stack & Development Approach: What tools, frameworks, or methods did you use to build it?

NeuroQC-Assist was developed mainly in Python and packaged as an Apptainer container so the workflow can be run reproducibly across computing environments. Image preparation uses neuroimaging tools, ANTs for registration-based processing and FSL utilities for field-of-view handling. The workflow standardizes T1-weighted MRI scans into a fixed 276 × 276 × 276 input size and generates corresponding brain mask. The image-quality model was built in PyTorch. It is a custom three-dimensional convolutional neural network, T1QCHybridCNN, that uses two input channels: the prepared T1-weighted image and the brain mask. The architecture uses an initial 3D convolutional stem followed by four residual-style convolutional stages, with skip connections and downsampling across feature-extraction layers. The network then uses global feature pooling and prediction heads for binary image-quality classification and artifact-type classification. The model was trained to estimate whether a scan is likely to be good or poor quality, with additional artifact labels used during development. Training data included visually acceptable scans, real poor-quality scans, and synthetic artifact examples created from good scans using data augmentation with TorchIO and MONAI. Synthetic artifacts included motion, ghosting, and Gibbs/ringing artifacts, were generated using transforms such as RandomMotion, RandomGhosting, and RandGibbsNoise. This was done to make the model see quality problems that may appear in real datasets. The workflow also includes Grad-CAM generation for visual review and spreadsheet for downstream use. The current development approach is practical and prototype-driven: build the end-to-end workflow first, test it on held-out scans, then refine thresholds, visualization, and deployment based on feedback.

Validation & Evidence: Have you done any informal testing or validation? If so, what did you find?

We have performed testing on held-out and external T1-weighted MRI scans to evaluate whether the workflow can separate clearly usable scans from scans that may need review or rejection. In the current prototype, the model produces a poor-quality confidence score for each scan. This score is converted into PASS, REVIEW, or FAIL categories. The REVIEW category is important because many scans are not obviously good or bad, and these borderline cases should remain in a human review workflow. On the validation split of 397 scans, the model achieved accuracy of 79.3% and an AUC of 0.784. Sensitivity, was 85.2%, Specificity, was 79.1%. On the test split of 143 scans from unseen datasets, the model correctly identified 35 of 45 poor-quality scans and 75 of 98 good-quality scans, corresponding to a sensitivity of 77.8% and a specificity of 76.5%. The model achieved an overall accuracy of 76.9%. Early testing suggests that the workflow can identify many poor-quality scans while still allowing clearly good scans to pass. It also showed that threshold choice matters: a stricter threshold can catch more problematic scans but may also send more acceptable scans to review. Because of this, we are treating the current thresholds as prototype settings rather than final validated cutoffs.

Feedback You're Seeking: What specific feedback, help, or collaboration are you hoping to get from the CAIMI26 community?

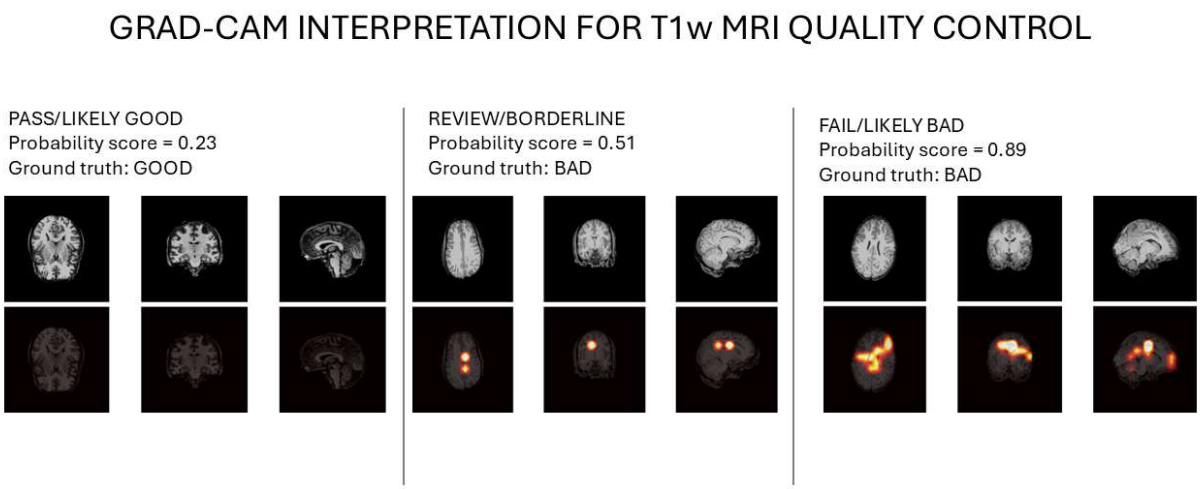
We would value feedback on where this approach could have clinical impact. Beyond research quality control, we are interested in whether the same framework could be adapted to protocol-specific quality checks in routine MRI workflows. We are also interested in identifying collaborators who could help us test these ideas on diverse datasets and shape the next phase of development around real clinical needs.

Known Limitations & Honest Failures: What isn't working yet, or what have you already tried that failed?

This is an early prototype and not a clinically validated tool. The model has been tested on limited datasets, and it needs broader evaluation across more scanners, institutions, acquisition protocols, and artifact types. The current thresholds are also not final. They are useful for demonstrating the PASS, REVIEW, and FAIL workflow, but they will need further tuning depending on the intended use case. Finally, the current MVP focuses on T1-weighted MRI only. The longer-term goal is a broader neuroimaging quality-control system, but other modalities will require separate development, testing, and validation.

Potential for Impact: If this MVP were fully developed and deployed, who would benefit and how?

If fully developed and validated, this Toolkit could help imaging teams review neuroimaging data more efficiently. Researchers could benefit by identifying problematic scans earlier, before time is spent on downstream modeling. Radiologists and clinical imaging teams could also benefit from a review-support tool that highlights scans needing closer inspection, while still keeping final judgment with a human reviewer. For multi-site studies, it could help make quality-control review more reproducible across datasets. The broader impact is not just automation, but better prioritization. The goal is to let clearly usable scans move forward, flag likely poor-quality scans, and keep uncertain cases in a human review pathway.



GitHub Repo:

https://github.com/SaiKrishnaChaitanya24/NeuroQC-AssistAI-Assisted-Neuroimaging-Quality-Control