NeuroMetrics: Automated Registration and Subtraction for Longitudinal Brain Imaging Assessment

Lucas I. Becker, MD MSc, University Medical Center Freiburg; Horst I. Urbach, MD

Briefly describe what attendees will see during your demo.

Attendees will see two serial brain imaging examinations imported into the prototype, automatically registered, and displayed with synchronized navigation. The demo will generate subtraction maps and quantitative change visualizations to highlight interval progression or regression. It will also show how the user can review alignment quality, inspect suspicious regions, compare measurements, and export clinically interpretable images and metrics for longitudinal assessment.

Problem & Motivation: What clinical or operational problem does your MVP address?

Longitudinal comparison of brain magnetic resonance imaging and computed tomography examinations is a routine task, which may be cognitively demanding. Small interval changes may be difficult to detect because of differences in patient positioning, slice orientation, acquisition parameters, and image contrast. Radiologists often rely on manual side-by-side comparison, visual memory, and repeated measurements, which are time-consuming and may be carried out by algorithms. This prototype was built to support a more reproducible longitudinal workflow. It automatically aligns serial examinations, enables synchronized navigation, produces subtraction-based change maps, and provides quantitative measurements of interval change. The objective is not to replace radiologist interpretation, but to reduce avoidable comparison burden and make subtle changes more visible and measurable. The initial use cases include follow-up of brain tumors, postoperative findings, and other focal intracranial abnormalities on magnetic resonance imaging or computed tomography. The tool is designed as an early-stage, locally deployable prototype that can be demonstrated on de-identified imaging data and refined through feedback from imaging informatics and clinical users.

What You Built: Describe your MVP, tool, or workflow in plain language

NeuroMetrics is a locally deployable prototype for longitudinal comparison of brain magnetic resonance imaging and computed tomography examinations. It accepts two or more serial imaging studies, converts and organizes the image data, aligns the follow-up examination to the baseline study, and presents both examinations in synchronized views. The tool generates subtraction maps and quantitative change visualizations intended to make interval progression or regression easier to identify. Users can review the registration result, scroll through both examinations simultaneously, inspect suspicious regions, and compare focal abnormalities across time. Depending on the use case, the workflow can also support lesion segmentation, volumetric comparison, and export of selected images or quantitative measurements. The initial applications include follow-up of brain tumors, cavernous malformations, postoperative findings, hemorrhagic lesions, and other focal intracranial abnormalities. Both magnetic resonance imaging and computed tomography can be processed, although performance depends on acquisition consistency and image quality. NeuroMetrics is not intended to provide an autonomous diagnosis. It is designed as a radiologist-facing comparison tool that reduces the manual effort of aligning examinations and makes subtle longitudinal change more visible and measurable.

Tech Stack & Development Approach: What tools, frameworks, or methods did you use to build it?
The prototype was built primarily in Python as a local, offline workflow for Windows-based clinical research use. Medical images are imported from Digital Imaging and Communications in Medicine format and converted when needed using dcm2niix. Image handling, resampling, registration, and subtraction rely on established medical imaging libraries, including SimpleITK, with Elastix-based registration tested for selected workflows. The interface was developed iteratively as a clinician-facing desktop prototype, with synchronized image navigation, registration review, subtraction display, and export functions. Three-dimensional Slicer has also been used for visual quality control, segmentation, and comparison with the prototype outputs. Development followed a rapid, clinician-led and partly “vibe-coded” approach: clinical needs were translated directly into small functional modules, tested on representative cases, and revised based on observed failure modes. Large language models were used to assist with code generation, debugging, interface design, and documentation, while all imaging logic, outputs, and clinical interpretations were reviewed manually. The current version runs locally without cloud processing or vendor-specific integration. It is an independent prototype rather than a commercial product or regulated medical device.

Validation & Evidence: Have you done any informal testing or validation? If so, what did you find?
Informal testing has been performed on selected longitudinal neuroimaging cases, including brain magnetic resonance imaging and computed tomography examinations with focal lesions, postoperative changes, hemorrhagic findings, and vascular malformations. The prototype has generally produced useful alignment and interpretable subtraction maps when baseline and follow-up examinations were acquired with comparable geometry, coverage, and image contrast. Synchronized navigation has been particularly helpful for reviewing small focal changes that are difficult to appreciate during conventional side-by-side comparison. Testing also showed that subtraction maps are highly sensitive to imperfect registration, differences in head position, acquisition parameters, motion, susceptibility artifacts, and postoperative anatomical distortion. Apparent interval change may therefore represent misregistration or contrast differences rather than true biological change. Manual review of registration quality remains mandatory. No formal accuracy, reader-performance, prospective clinical, or outcome validation has yet been completed. The current evidence is therefore limited to feasibility, technical testing, and developer use on representative cases. These observations have been useful for identifying both promising use cases and the conditions under which the workflow becomes unreliable.

Feedback You're Seeking: What specific feedback, help, or collaboration are you hoping to get from the CAIMI26 community?
I am seeking feedback on how to transform a functional radiologist-built prototype into a robust and clinically evaluable longitudinal imaging tool. Specific questions include how best to validate registration and subtraction performance across scanners, institutions, sequences, and disease entities; which quantitative endpoints would be most clinically meaningful; and how to design a reader study that measures whether the tool improves sensitivity, reporting time, or confidence without increasing false-positive findings. I would also value input on user-interface design, automated quality control, uncertainty visualization, and methods for distinguishing true interval change from registration or acquisition artifacts. Advice on integration into picture archiving and communication systems, deployment within hospital networks, and handling of Digital Imaging and Communications in Medicine metadata would be particularly helpful. I am interested in collaboration with imaging informaticians, neuroradiologists, computer vision researchers, and institutions willing to support

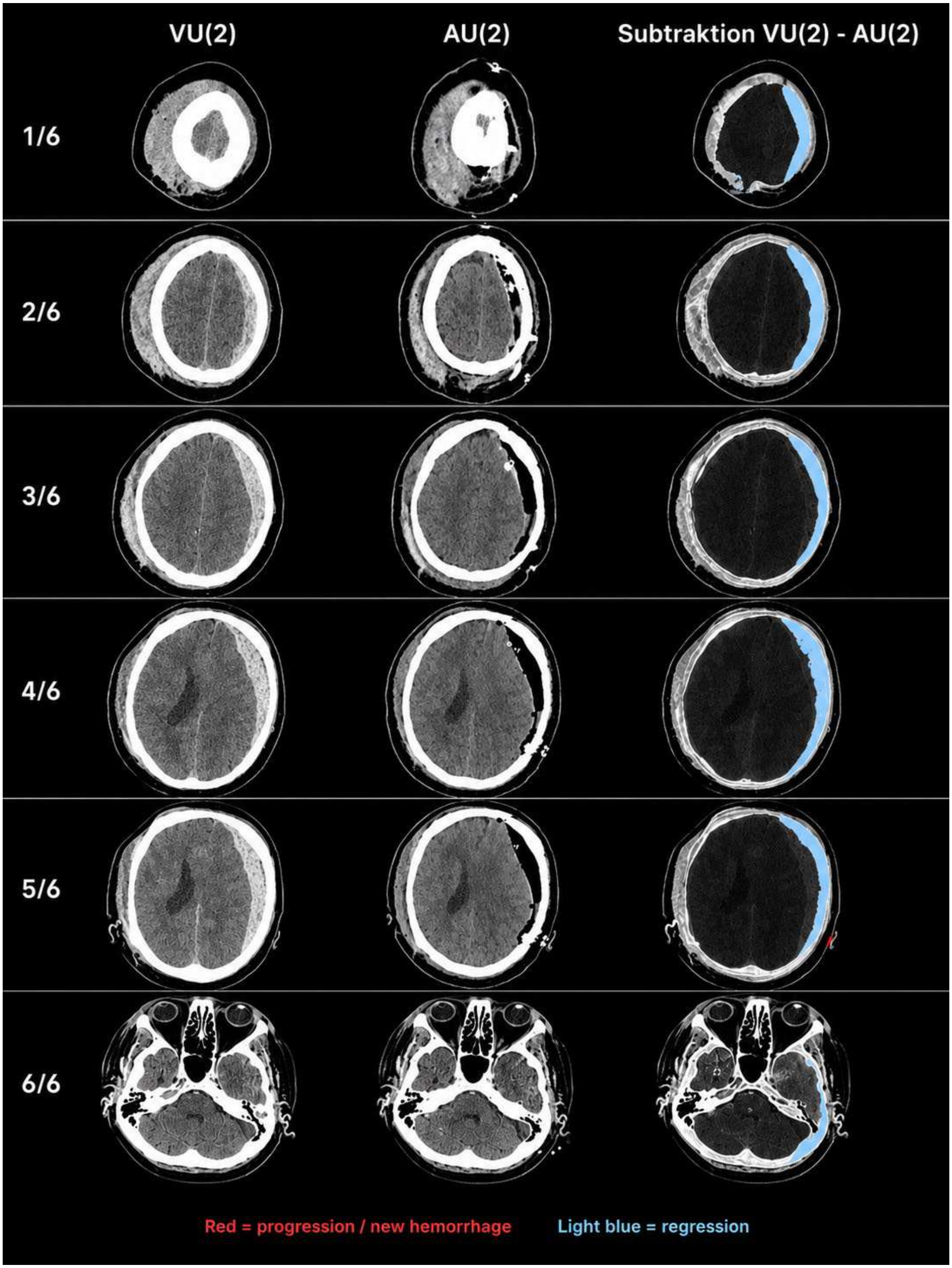
external testing. The most valuable outcome of the session would be a clearer validation pathway, identification of priority use cases, and guidance on which technical features are essential before broader clinical evaluation.

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

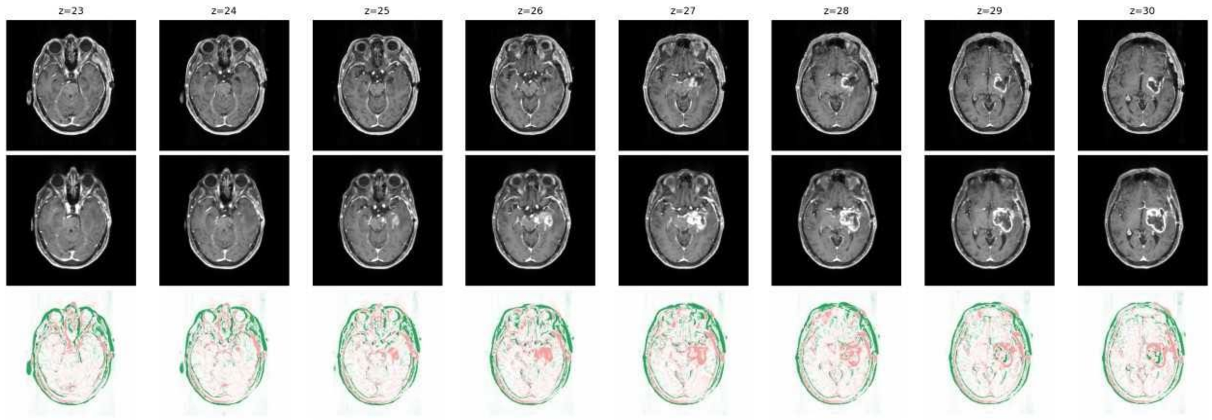
The main limitation is that subtraction quality depends heavily on accurate registration and sufficiently comparable source examinations. Differences in slice thickness, angulation, field of view, contrast timing, magnetic resonance imaging sequence parameters, motion, susceptibility artifacts, and postoperative deformation can generate misleading subtraction signals, which is why I opted for rigid registration with 1 mm Slice thickness images. Rigid registration is often insufficient for complex anatomical change, while deformable registration may improve alignment but can also distort pathology and create false impressions of progression or regression. Automated registration therefore cannot currently be trusted without visual quality control. The prototype does not yet provide validated automated registration-failure detection, calibrated uncertainty estimates, robust lesion tracking across all modalities, or standardized quantitative thresholds for clinically meaningful change. Some early registration approaches failed because slice orientation and spacing were not handled consistently. Two-dimensional registration was also inadequate when examinations differed substantially in three-dimensional geometry. The current workflow is not integrated with a clinical picture archiving and communication system or reporting platform. Automated query and retrieval, structured export, user authentication, audit trails, and scalable deployment would be difficult without institutional information technology support and, in some environments, vendor interfaces. The tool is not formally validated, regulated, or intended for unsupervised clinical decision-making.

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

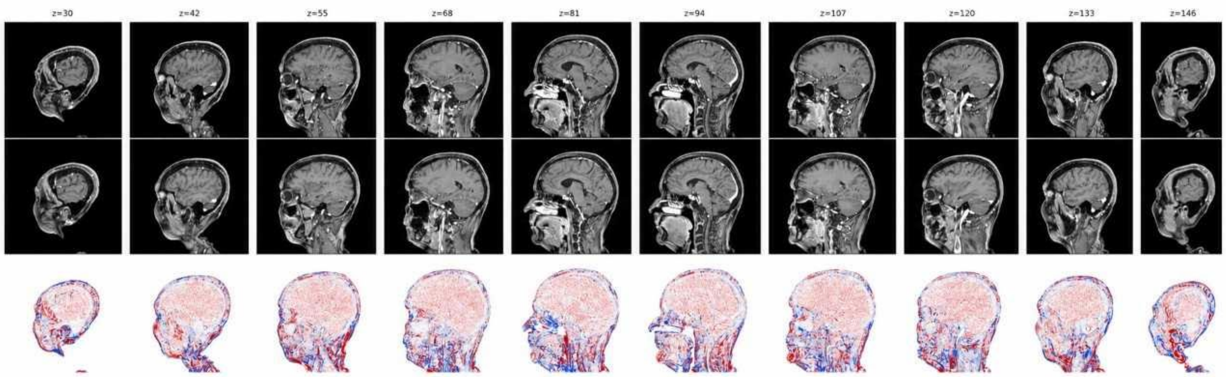
If fully developed and validated, NeuroMetrics could support radiologists by reducing the time and cognitive effort required for longitudinal comparison of brain imaging examinations. Registration, synchronized navigation, subtraction maps, and quantitative measurements could help make small interval changes more visible, particularly in complex follow-up studies with multiple lesions or subtle progression. Patients could benefit through earlier recognition of treatment response, tumor progression, recurrent hemorrhage, postoperative change, or other clinically relevant abnormalities. More reproducible comparison may also reduce unnecessary repeat imaging and improve confidence in multidisciplinary treatment decisions. Trainees could use the tool to learn how lesions evolve over time and to understand the limitations of visual comparison and image registration. Researchers could use standardized longitudinal outputs for segmentation, volumetry, radiomics, and treatment-response studies. At the institutional level, a local and vendor-neutral workflow could provide a flexible platform for evaluating longitudinal imaging methods without requiring immediate commercial deployment. The broader goal is not autonomous interpretation, but a transparent and controllable augmentation tool that helps radiologists compare prior and current examinations more consistently, efficiently, and safely.


VU(2)
AU(2)
Subtraktion VU(2) - AU(2)
1/6
2/6
3/6
4/6
5/6
6/6
Red = progression / new hemorrhage
Light blue = regression

MPRAGE Verlauf: VU vs. AU und Differenzkarten (Zunahme = hellrot, Abnahme = grün)



MPRAGE post Gd Kopf | VU 04.03.2026 vs. AU 09.06.2026



Keine Tumordetektion

Neurometrics (Becker, Chen, 2026)