Original Paper

Magnetic Resonance Imaging–Based Measurement of Pediatric Head Circumference by a Clinician-in-the-Loop 3D Surface Tool: Development and Validation Study

Leonie Dittrich[1]; Jan Philipp Adam[2]; Ute von Jan[1]; Eva Bültmann[2§]; Dominik Wolff[1§]

[1]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover Medical School, Hannover, Lower Saxony, Germany

[2]Institute of Diagnostic and Interventional Neuroradiology, Hannover Medical School, Hannover, Lower Saxony, Germany

§These authors share senior authorship.

Corresponding Author:

Leonie Dittrich
Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School
Hannover Medical School
Carl-Neuberg-Straße 1
Hannover, Lower Saxony 30625
Germany
Phone: 49 5115325372
Email: leonie.dittrich@plri.de

Abstract

Background: Head circumference (HC) is a routine clinical parameter in pediatric patients indicating brain growth and neurodevelopment. Although usually obtained with a tape measure, this method is operator-dependent and prone to variability. Magnetic resonance imaging (MRI) provides a retrospective alternative, but existing imaging-based approaches can be limited by predefined measurement planes, restricted clinician review, and cranial surface artifacts.

Objective: This study aimed to develop and evaluate a clinician-guided software tool for measuring pediatric HC from routinely acquired 3D MRI examinations.

Methods: We developed a desktop application in Python using PySide6, the Insight Segmentation and Registration Toolkit, and the Visualization Toolkit for MRI-based HC measurement. The workflow included outer skin surface reconstruction, placement of 3 landmarks to define a freely orientable measurement plane, contour extraction, and circumference calculation. Optional Delaunay-based interpolation was implemented to reduce the influence of local cranial surface irregularities. In this retrospective evaluation, the tool was applied to 23 pediatric MRI examinations, including 18 without and 5 with clinically identified cranial surface irregularities. Two clinicians independently performed 3 repeated software measurements per examination. Median software measurements were compared with a single tape-based HC measurement obtained on the date of MRI. Agreement was assessed using Bland-Altman analysis, mean absolute error (MAE), proportional bias analysis, and a linear mixed-effects model. Interrater differences and preliminary usability feedback were also evaluated.

Results: Software-based HC measurement was feasible in all 23 examinations. In cases without cranial surface irregularities, the raw and interpolated contours showed similar agreement with tape-based measurements (bias 0.415 cm and 0.410 cm; MAE 0.429 cm and 0.423 cm, respectively). In the 5 examinations with irregularities, the intersection contour showed a greater bias of 1.420 cm and wider limits of agreement (LoA; –0.182 to 3.021 cm), whereas the interpolated contour showed a lower bias of 0.448 cm and narrower LoA (–0.119 to 1.014 cm). The estimated coefficient for the contour-by-subgroup interaction was –0.967 cm (SE 0.195 cm, 95% CI –1.349 to –0.584 cm; $P<.001$). No statistically significant proportional trend was detected, although the analysis had limited power to exclude proportional bias. Interrater differences were small. The 2 clinicians reported System Usability Scale scores of 92.5 and 95.0, representing preliminary descriptive feedback.

Conclusions: The clinician-guided workflow enabled MRI-based HC measurement in all examined cases and showed small interrater differences. Delaunay-based interpolation was associated with closer agreement with tape-based measurements in the small subgroup with cranial surface irregularities, while both contour variants performed similarly in examinations without irregularities. These findings are preliminary because of the small cohort, the single tape-based reference measurement per

patient, and the 5-case irregularity subgroup. Larger studies using repeated reference measurements and broader cranial morphologies are needed to validate the workflow.

JMIR Med Inform 2026;14:e97692; doi: 10.2196/97692

Keywords: head circumference; magnetic resonance imaging; pediatric MRI; cranial morphometry; medical image analysis; neuropediatrics; 3D surface reconstruction; image-based measurement

Introduction

Head circumference (HC) is a common pediatric anthropometric measurement and an important indicator of brain growth and neurodevelopment. HC measurements outside age- and sex-specific reference ranges may reflect altered cranial or brain growth and can be associated with a range of neurological, developmental, or genetic conditions and warrant further clinical evaluation [1-5]. Because clinicians track HC throughout infancy and childhood, measurement variability can affect percentile assignment and clinical interpretation, especially near decision thresholds [1].

In clinical practice, HC is usually measured with a flexible tape placed around the largest circumference of the head, passing in front over the supraorbital region and behind the most prominent occipital point. Although inexpensive and widely available, tape-based HC measurement is operator-dependent and sensitive to head positioning, tape tension/stretchability, hair or scalp compression, and patient compliance. This leads to clinically relevant intra- and interobserver variability [6-8]. Standardized procedures and nonstretchable tapes can improve reproducibility, but measurements remain technique-dependent and require training [6]. Furthermore, manual tape measurements are often unavailable for retrospective studies that rely on imaging archives and can be difficult to obtain from children who cooperate poorly or have atypical cranial morphology [9].

When magnetic resonance imaging (MRI) data are available, imaging-based HC estimation allows retrospective measurement of HC at the time of image acquisition. Available approaches differ in how they define the measurement plane and outer head contour and in the degree of clinician involvement. Vorperian et al [10] estimated HC indirectly from 2 imaging-derived cranial dimensions using growth-adjusted geometric models, avoiding full surface segmentation but potentially limiting applicability in atypical head shapes. Rau et al [11] used multiplanar MRI reformations with user-adjusted axial plane selection and an ovoid region of interest to approximate the outer head contour, providing case-specific control at the cost of operator dependence. Nevertheless, the method showed an excellent interrater agreement. Fully automated approaches reduce reader interaction but rely strongly on successful segmentation and predefined plane-selection rules. Yepes-Calderón et al [12] identified the maximum axial circumference within a Picture Archiving and Communication System (PACS)-integrated MRI workflow, supporting efficient integration into clinical reporting while limiting clinical control, for example, in cases with atypical anatomy. Acamine et al [13] applied a similar maximum-slice principle to computed tomography data, including scans from children with head abnormalities. The method enables automated retrospective measurement but remains dependent on axial sampling and accurate contour detection. Outside radiological imaging, Beaumont et al [14] used 3 anatomical landmarks to define a measurement plane on an externally acquired 3D head surface. This approach is geometrically similar to the current workflow but requires a dedicated surface scan and therefore does not allow retrospective measurement from routine MRI data.

More recently, Adam et al [15] evaluated manual, semiautomated, and fully automated MRI-based combinations of visual-, rule-, atlas-, and neural network-based plane selection with ellipsoid- or contour-based circumference estimation using the same cohort and tape-based reference measurements as this study. Manual methods allowed case-specific adjustment but were reader-dependent and time-consuming, whereas automated methods improved scalability but relied on successful segmentation. Their comparison indicated that manual methods and automated ellipsoid approaches achieved similarly acceptable agreement with tape-based HC, while automated contour methods were less accurate and more susceptible to segmentation-related overestimation. Klischat and Wilke [16] proposed a fully automated landmark-transfer and convex-hull approach across tilted planes. This supports efficient processing of larger datasets but offers limited case-specific adjustment and may be less robust in severely abnormal head configurations.

Overall, existing methods rely on indirect geometric estimation, user-defined or automatically determined measurement planes, maximum-circumference selection from axial or tilted slices, or dedicated 3D surface acquisition [10-16]. Their performance may be affected by head position, image orientation, slice geometry, segmentation quality, and landmark transfer [11-16]. Geometric approximations and automated contour methods may also be less robust in atypical or locally distorted head shapes, while surface-scanning approaches require a separate acquisition and do not allow retrospective measurement from routine MRI data [10, 14-16].

To address these limitations, we developed a clinician-guided tool for measuring HC from 3D pediatric MRI data [17]. Rather than relying on predefined, automatically transferred, or slice-based measurement planes, the tool reconstructs the outer head surface and allows the user to define a freely orientable plane by placing 3 landmarks directly on the surface. The user can visually review the reconstructed surface, landmark positions, measurement plane, and resulting contour before accepting the measurement. The software then computes HC from the intersection contour

and provides optional Delaunay-based interpolation to reduce the influence of local surface irregularities. In this study, we evaluated the feasibility of this workflow, its agreement with same-day tape-based HC measurements, interrater agreement based on repeated measurements by 2 clinicians (EB and JPA), and preliminary usability.

Methods

Ethical Considerations

The study has been performed in accordance with the Declaration of Helsinki and was approved by the ethics committee of Hannover Medical School (number 12478-BO-K-2026). Ethical approval for the secondary analysis of the previously acquired imaging and clinical data was obtained retrospectively in 2026 and covers the analyses reported in this study. All subjects' parents provided informed consent for secondary use of their data. Before analysis, direct patient identifiers were removed, and the data were processed using study-specific identifiers only. No compensation was provided to participants in connection with this study.

Software Design and Implementation

Overview

A dedicated desktop application was developed to enable interactive HC measurement from 3D surface reconstructions derived from pediatric MRI. The software was implemented in Python (v3.12) using Qt (PySide6) for the graphical user interface and Insight Segmentation and Registration Toolkit (ITK), and the Visualization Toolkit (VTK) for medical image input/output, visualization, and geometric processing. The application supports Digital Imaging and Communications in Medicine (DICOM) input as well as volumetric formats, including .nrrd, .vtk, and .vtu. While the workflow is imaging-sequence-agnostic in principle (requiring a 3D volume with sufficient contrast for outer head surface extraction), it was primarily designed for and applied to T1-weighted 3D magnetization-prepared rapid acquisition gradient echo (MPRAGE) data. The workflow includes surface reconstruction, landmark-defined plane selection, contour generation, and circumference calculation, with an optional contour interpolation step.

Outer Skin Surface Reconstruction

The measurements were performed on a 3D surface intended to represent the outer skin boundary. The surface was extracted from the raw MRI intensity volume after median filtering (5×5×5 voxels; vtkImageMedian3D) using isosurface extraction at intensity t (vtkContourFilter). To minimize stair-step artifacts from voxel discretization, the resulting mesh was smoothed using vtkSmoothPolyDataFilter (30 iterations, relaxation factor 0.5). Any disconnected fragments were removed by retaining only the largest connected component (vtkConnectivityFilter, largest region). Despite using imaging data from 4 different scanners at either 1.5 or 3 Tesla (T) in the study (Table 1), a consistent threshold of t=20 (MRI intensity units) was applied to all datasets. This threshold was selected based on preliminary testing in an MRI dataset not included in the study cohort. To isolate the effect of the threshold, identical landmark coordinates and therefore an identical measurement plane were used for all tested threshold values. The landmark and threshold configuration could be saved and reloaded within the application, allowing the same cut-plane geometry to be applied across all threshold settings. The resulting circumferences were 536.879 mm at t=10, 533.894 mm at t=15, 532.399 mm at t=20, 531.400 mm at t=25, and 530.590 mm at t=30. Lowering the threshold to t=10 additionally resulted in visible artifacts around the nose and neck. Because this sensitivity analysis was performed on a single dataset outside the study cohort, it was intended as a preliminary assessment and does not establish threshold robustness across scanners or acquisition protocols.

Table 1. Demographic characteristics, head circumference, clinical indication, and MRI[a] scanner distribution of the study cohort.

Characteristic	Value
MRI examinations, n	• 23
Age at MRI, months	• Median 49.2 (IQR 25.7-79.8) • Range 13.2-206.6
Sex, n (%)	• Male: 14 (60.9%) • Female: 9 (39.1%)
Tape-based HC[b], cm	• Median 48.90 (IQR 45.85-52.70) • Range 38.00-54.60
Clinical indication, n	• Sensorineural hearing loss: 10 • Tumor follow-up: 5 • Exclusion of intracranial tumor: 2 • Intracranial aneurysm: 2 • Posttraumatic findings: 1 • Hydrocephalus follow-up: 1 • Suspected encephalitis: 1 • Developmental delay: 1
MRI scanner, n	• 3 T[c] Verio: 11 • 1.5 T Aera: 10

Characteristic	Value
	• 1.5 T Avanto: 1 • 3 T Skyra: 1

[a]MRI: magnetic resonance imaging.
[b]HC: head circumference.
[c]T: Tesla.

All T1-weighted volumes were acquired with the Prescan normalize feature enabled, which corrects for spatial intensity variation related to receiver-coil sensitivity. A fixed threshold of t=20 was used as a practical approach across all datasets. Because MRI intensity values are not inherently standardized across scanners or field strengths, every generated surface was visually inspected in the 3D viewer before measurement. Although the software allowed manual adjustment of the threshold when necessary, no threshold adjustments were required for any dataset in the study cohort.

Landmark-Defined Measurement Plane and Contour Generation

HC was computed as the perimeter of the intersection between the reconstructed surface and a user-defined plane. To obtain the parameters for this plane, the users selected 3 points, P1, P2, and P3, on the 3D surface that approximated the positions that would also have been in the areas touched by the tape measure using vtkCellPicker. These points were used to define a unique landmark-defined measurement plane in point-normal form: P1 served as the plane's origin, and 2 in-plane vectors were computed as V1=P2−P1 and V2=P3−P1. The plane normal was then obtained via the cross-product N=V1×V2. The plane was instantiated as a vtkPlane using P1 and N, and the intersection contour between the plane and the surface mesh was generated using vtkPlaneCutter, producing a polyline representation of the cross-section. In cases of suboptimal point placement, the tool allows deletion and replacement of interactively defined landmarks until an anatomically acceptable plane and contour are obtained.

Circumference Calculation and Optional Contour Interpolation

The circumference was computed from the intersection polyline using vtkIntegrateAttributes, which provides polyline length in physical units. The resulting value was reported in millimeters and converted to centimeters for clinical interpretation. The accuracy and numerical precision of the circumference calculation were assessed using a purely software-based cylindrical phantom. Two radii were evaluated: r=3 mm (n=64 facets) and r=70 mm (n=360 facets), with an expected circumference $C=2\pi r$ of approximately 18.85 mm and 439.82 mm, respectively. A total of 10 independent measurements were performed on the equatorial plane of each cylinder. For r=3 mm, the measured values ranged from 18.841 to 18.844 mm (mean 18.842, SD 0.0009), while for r=70 mm, they ranged from 439.817 to 439.821 mm (mean 439.8183, SD 0.0013). The slightly lower values compared with the theoretical circumferences are consistent with the perimeter of an inscribed n-sided polygon being smaller than the circumference of the corresponding circle. Because the landmarks were placed on the equatorial plane of an idealized cylindrical surface, this experiment was not designed to assess variability in landmark placement.

To address local surface irregularities (eg, artifacts due to shunt devices) that can produce jagged contours, the software optionally generates an interpolated contour. For this, the original contour points are converted into a point cloud (vtkConvertToPointCloud) and used to compute a two-dimensional (2D) Delaunay triangulation (vtkDelaunay2D) with best-fitting plane projection. The boundary edges of the triangulated mesh are then extracted (vtkFeatureEdges) to form an interpolated polygonal contour. Both the original and interpolated contours are visualized. No artificial surface defects or irregularities were introduced into the cylindrical phantom. Therefore, the phantom experiment verifies the circumference calculation but does not validate the ability of Delaunay interpolation to mitigate irregular surface contours.

Software Environment

Software library versions and execution environment are reported as follows: Python 3.12 (Python Software Foundation), VTK v9.3.1 (Kitware Inc), ITK v5.4.0 (Insight Software Consortium), PySide6 v6.7.2 (The Qt Company); operating systems Windows 11 23H2 (Microsoft Corporation; build 22631.6491) and macOS (Apple Inc; MacBook Air M3 2024, 24 GB RAM, macOS Tahoe 26.3).

Clinical Evaluation

Study Design and Cohort

A clinical evaluation was conducted to compare software-derived HC against tape-based HC measurement, which represents the current gold standard. The cohort comprised 23 pediatric head MRI examinations acquired at Hannover Medical School from January to March 2024. Demographic, clinical, and imaging characteristics of the study cohort are summarized in Table 1. All examinations performed during this period were deliberately included in the study in order to validate the software under real-world conditions, including the presence of artifacts caused, for example, by shunt devices.

The MRI examinations and tape-based reference measurements included in this study were also used in a different study by JPA and EB [15]. The current study evaluates a separate clinician-guided software workflow, and all software development, software-derived measurements, and statistical analyses reported here were generated specifically for this study.

MRI Acquisition

All MRI studies included a sagittal 3D-T1w-MPRAGE with 1 mm slice thickness, which was used for this study. Acquisition was done using different MRI scanners (Siemens 1,5T Aera and Avanto as well as 3T Verio and Skyra, Siemens Healthineers) with the following parameters: repetition time 2200/1980/1800/2300 ms, echo time 2.67/3.98/2.27/2.32 ms, and flip angle 8°/15°/9°/8° (Aera/Avanto/Verio/Skyra). Imaging data were exported from the PACS at Hannover Medical School in DICOM format and processed within the software.

Tape-Based HC Measurement

Tape-based HC was measured in centimeters using a nonstretchable measuring tape placed around the maximal cranial circumference, passing anteriorly over the supraorbital region and posteriorly over the most prominent occipital point. Tape measurements were obtained by EB on the day of MRI acquisition and served as the reference for the agreement analysis. Further details on the tape-based measurement procedure are provided by Adam et al [15]. No repeated tape measurements were available for this study.

Software-Based Measurement Procedure

For the software-based measurements, 2 clinicians (JPA and EB) independently measured HC for all MRI examinations without access to each other's results. For each case, the dataset was loaded, and the reconstructed outer skin surface was verified. The measurement plane was defined by placing 3 landmarks directly on the outer surface: 2 anterior points at the most prominent points of the forehead on the right and left sides, and one posterior point at the most prominent point of the occiput. The plane was then visually adjusted to ensure that the circumference passed above the ear and to avoid corresponding distortions. The resulting surface-plane intersection contour was visually confirmed, and HC was recorded in millimeters.

Cranial surface irregularities were classified by EB based on the visual assessment of the reconstructed outer head surface. Five cases were classified as having cranial surface irregularities, of which 2 involved a shunt device, 2 showed occipital flattening, and one showed a local distortion caused by an external cable. The irregularity classification was performed before reviewing the results of either contour method, thereby preventing the classification from being influenced by the subsequent performance of one approach. Independently of this classification, both the direct intersection contour and the Delaunay-interpolated contour were calculated for all examinations to enable direct comparison of the 2 approaches, while the latter was intended as a regularized alternative for cases with cranial surface irregularities.

Each clinician repeated the software-based measurement 3 times to assess repeatability and account for variability in manual landmark placement. Both clinicians were blinded to the documented tape-based reference values during the software-based measurements. However, because EB had originally performed the tape-based measurements, recall of individual measurements cannot be completely excluded, despite an interval of approximately 2 years between the tape-based and software-based measurements.

Usability Assessment

Usability was assessed using the System Usability Scale (SUS). After completing the measurement tasks, the two participating clinicians independently answered the 10-item SUS questionnaire on a 5-point Likert scale. Scores were computed using the standard procedure and reported on a 0-100 scale, where higher values indicate better-perceived usability.

Statistical Analysis

Statistical analyses were conducted in a separate Python 3.13 environment, using pandas 2.3.3, NumPy 2.4.1, SciPy 1.17.0, and statsmodels 0.14.6, independently of the software development, which used Python 3.12. Agreement between software-derived and tape-based HC was evaluated at the rater level. Each clinician performed 3 repeated measurements per head for each contour variant; repetitions were aggregated using the median within each head-clinician-contour combination, resulting in 92 observations (23 heads×2 clinicians×2 contour types). Tape-based HC was available as a single value per subject and was used as the reference.

Agreement was quantified using a Bland-Altman analysis, with the difference defined as (software measurement–tape measurement) and the mean of the 2 methods defined as $\left(\frac{mean\ software\ measurement + tape\ measurement}{2}\right)$. Bias and 95% LoA (bias±1.96×SD of differences) were computed separately for each contour and stratified by subgroup (no cranial surface irregularity vs cranial surface irregularity). Error magnitude was summarized by mean absolute error (MAE).

Uncertainty for the Bland-Altman metrics (bias, LoA, and MAE) was quantified using cluster bootstrap 95% CI with resampling by head ID. A linear mixed-effects model was fitted to the signed difference between software-derived and tape-based HC. Contour type, cranial surface irregularity subgroup, their interaction, and clinician were included as fixed effects. To account for repeated observations within patients and contour type as a within-patient factor, the model included a patient-specific random intercept and a patient-specific random slope for contour type. Proportional bias was assessed using Bland-Altman trend analyses (difference vs mean) for each contour. Interrater agreement between clinicians was assessed using Bland-Altman analysis on per-head medians (Clinician 1–Clinician 2).

Results

Cohort and Feasibility

The workflow produced valid HC measurements in all 23 examinations. Both the intersection contour and the Delaunay-interpolated contour were evaluated for all cases to enable comparison. Five cases were classified as having

cranial surface irregularities. Figures 1 and 2 illustrate the measurement workflow displayed in the software interface. Figure 1 shows a case without irregularities, for which the raw plane-surface intersection was sufficient. Figure 2 shows an irregular case with an inward-directed surface reconstruction due to a shunt artifact and illustrates the Delaunay-interpolated contour.

Figure 1. Representative example of a head circumference measurement in a case without cranial surface irregularities. The reconstructed outer head surface is displayed in anterior (top), lateral (middle), and posterior (bottom) views. Black markers denote the 3 clinician-selected landmarks that define the measurement plane. The white contour represents the raw plane-surface intersection from which the circumference is calculated. Surface detail below the measurement contour has been grayed out to preserve patient privacy.

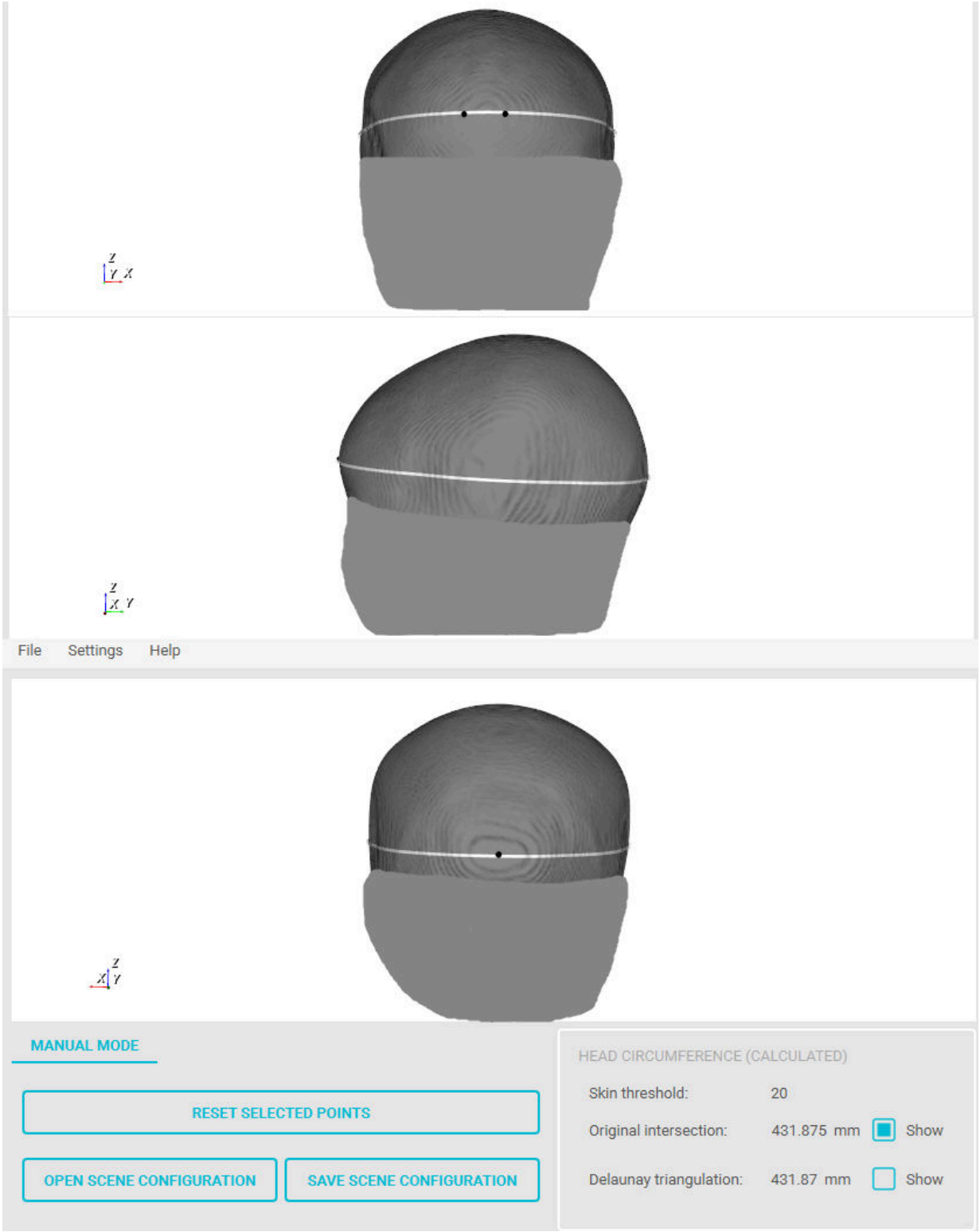
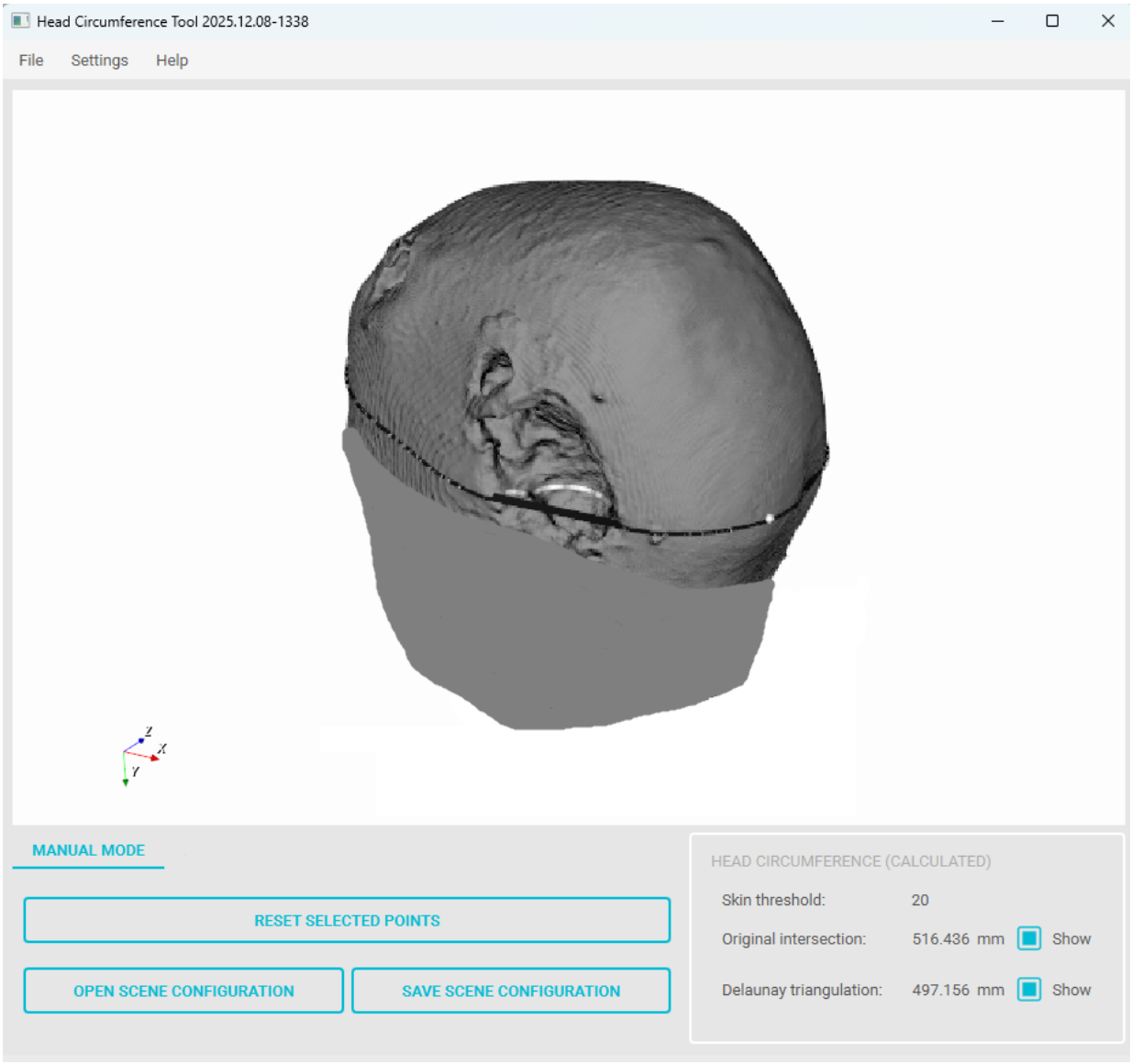


Figure 2. Representative example of a head circumference measurement in a case with a cranial surface irregularity (inward-directed surface due to shunt artifact). The reconstructed outer head surface is displayed in a lateral view. White markers denote the 3 clinician-selected landmarks that define the measurement plane. The white contour shows the raw plane-surface intersection, whereas the black contour shows the Delaunay-interpolated contour. Surface detail below the measurement contour has been grayed out to preserve patient privacy.



Agreement With Tape-Based HC

We assessed agreement between software-derived and tape-based HC using Bland-Altman analyses on per-head medians of 3 repetitions per clinician. Among the 18 patients without cranial surface irregularities (36 rater-level pairs per contour), both contour variants produced similar biases (Delaunay: 0.410 cm; intersection: 0.415 cm) and error magnitudes (MAE: 0.423 cm vs 0.429 cm; Table 2). The 95% LoA were comparable between the 2 variants: intersection, –0.313 to 1.143 cm; Delaunay, –0.305 to 1.125 cm (Table 2). Figure 3 visualizes all paired observations and shows that larger positive differences occurred predominantly in the cranial surface irregularity subgroup for the raw intersection contour. In contrast, measurements based on the Delaunay-interpolated contour showed a narrower distribution of differences.

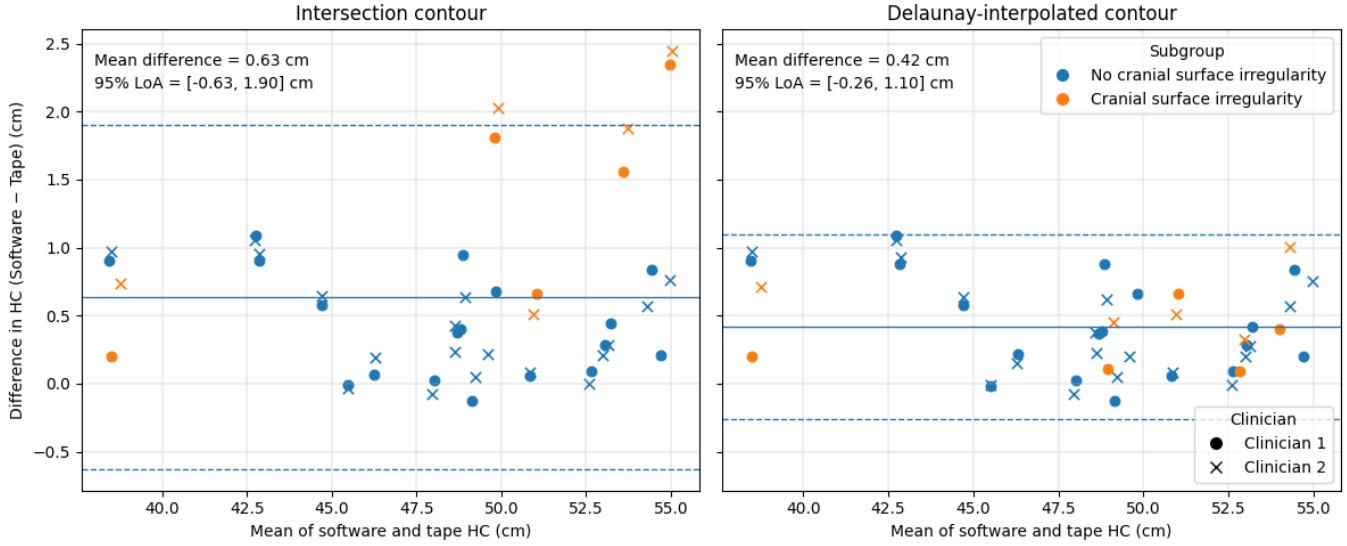
Table 2. Bland-Altman agreement between software-derived and tape-based head circumference at the rater level, stratified by subgroup.[a]

Contour and subgroup	Pairs, n	Bias (software–tape), cm	95% LoA[b] (low), cm	95% LoA[b] (high), cm	MAE[c], cm
Intersection contour					
No cranial surface irregularity	36	0.415 (0.262 to 0.578)	–0.313 (–0.448 to -0.100)	1.143 (0.849 to 1.355)	0.429 (0.284 to 0.585)
Cranial surface irregularity	10	1.420 (0.768 to 2.071)	–0.182 (–0.681 to 1.459)	3.021 (1.856 to 3.394)	1.420 (0.768 to 2.071)
Delaunay-interpolated contour					

Contour and subgroup	Pairs, n	Bias (software–tape), cm	95% LoA[b] (low), cm	95% LoA[b] (high), cm	MAE[c], cm
No cranial surface irregularity	36	0.410 (0.261 to 0.569)	–0.305 (–0.438 to -0.102)	1.125 (0.833 to 1.342)	0.423 (0.283 to 0.575)
Cranial surface irregularity	10	0.448 (0.284 to 0.609)	–0.119 (–0.223 to 0.104)	1.014 (0.675 to 1.217)	0.448 (0.284 to 0.609)

[a]For each head and clinician, software measurements were aggregated as the median of 3 repetitions. Values are reported as point estimates with 95% cluster-bootstrap CIs (resampling patient ID). Differences are defined as (software–tape).
[b]LoA: limits of agreement.
[c]MAE: mean absolute error.

Figure 3. Bland-Altman plots comparing software-derived and tape-based head circumference (HC) measurements. Each point represents 1 clinician-specific comparison for 1 head and contour type, where the software value was defined as the median of 3 repeated measurements. For each of these comparisons, the x-axis was calculated as the mean of the clinician-specific software value and the tape measurement, that is ($\frac{software\ mean+tape}{2}$) and, the y-axis as their difference (software–tape). Solid lines indicate the mean difference, and dashed lines indicate the 95% limits of agreement. The inset mean differences of 0.63 cm for the intersection contour and 0.42 cm for the Delaunay-interpolated contour refer to the pooled cohort across all 23 examinations and not to the subgroup-specific values. HC: Head circumference; LOA: limits of agreement.



In patients with cranial surface irregularities (5 heads; 10 pairs per contour), the intersection contour yielded a larger bias (1.420 cm) and wider LoA (upper LoA: 3.021 cm), whereas the Delaunay-interpolated contour yielded a bias of 0.448 cm with an upper LoA of 1.014 cm (Table 2). We evaluated proportional bias by regressing the difference between methods on the mean of the 2 measurements (difference~mean). The estimated slopes were 0.017 cm for the intersection contour (P=.55) and –0.023 cm for the Delaunay-interpolated contour (P=.10), and neither was statistically significant. However, the limited sample size restricts the ability to rule out proportional bias. The linear mixed-effects model included 92 aggregated measurements from 23 heads and converged successfully. Using the Delaunay contour and the irregularity subgroup as reference categories, the contour-by-subgroup interaction was significant (estimate –0.967 cm, SE 0.195, 95% CI –1.349 to –0.584 cm; P<.001), whereas the estimated difference in heads without irregularities was 0.005 cm.

Interrater Agreement

The interrater agreement between Clinicians 1 and 2 was evaluated using the median of 3 repetitions per head (n=23). The mean difference between Clinician 1 and Clinician 2 was –0.020 cm for the intersection contour and –0.036 cm for the Delaunay-interpolated contour. The 95% LoA were –0.533 to 0.492 cm for the intersection contour, and –0.597 to 0.525 cm for the Delaunay-interpolated contour (Table 3).

Table 3. Interrater agreement between clinician 1 and clinician 2 for software-derived head circumference[a].

Contour	Heads, n	Mean difference (C1–C2), cm	95% LoA[b] (low), cm	95% LoA (high), cm	MAE[c], cm
Intersection contour	23	–0.020	–0.533	0.492	0.195
Delaunay-interpolated contour	23	–0.036	–0.597	0.525	0.213

[a]Values are based on per-head medians of 3 repeated measurements per clinician (n=23 heads). Differences are reported as (Clinician 1–Clinician 2).
[b]LoA: limits of agreement.
[c]MAE: mean absolute error.

Usability

Usability was assessed by 2 clinicians (JPA and EB) using the SUS. The SUS scores were 92.5 and 95.0.

Discussion

Principal Findings

This study evaluated a user-guided workflow for measuring HC from 3D pediatric MRI scans. This workflow consisted of outer-surface reconstruction, landmark-defined plane selection, and contour-based circumference computation. The approach was feasible in this cohort and produced valid measurements in all 23 examinations. Although both contour variants were evaluated in all examinations for direct comparison, the Delaunay-interpolated contour was specifically intended as a regularized alternative for cases with cranial surface irregularities. Agreement with tape-based HC depends on the subgroup and contour variant. Among heads without cranial surface irregularities (n=18; 36 rater-level pairs per contour), both contour variants exhibited nearly identical mean differences of approximately 0.41 cm (intersection: 0.415 cm; Delaunay interpolated: 0.410 cm), with similar 95% LoA (–0.313 to 1.143 cm and –0.305 to 1.125 cm), as well as comparable error magnitudes (MAE ~0.42 to 0.43 cm). These findings suggest that interpolation did not affect performance in typical cases.

In the irregularity subgroup (5 heads; 10 pairs per contour), the intersection contour showed a pronounced increase in bias and dispersion (mean difference: 1.420 cm; upper LoA: 3.021 cm). In contrast, the Delaunay-interpolated contour remained closer to the tape (mean difference: 0.448 cm; upper LoA: 1.014 cm). This is consistent with the idea that interpolation mitigates irregularity-driven contour excursions. The mixed-effects model showed a significant contour-by-subgroup interaction (P<.001), indicating that the difference between the 2 contour variants was substantially greater in heads with cranial surface irregularities. The model converged successfully, and the interaction remained significant after accounting for patient-specific random slopes for contour type. However, the interaction estimate should be interpreted cautiously because the irregularity subgroup was small. The limited number and range of irregular cranial morphologies restrict the stability and generalizability of this finding.

Proportional-bias analyses showed small, nonsignificant slopes for both contour variants. However, the nonsignificant slopes should not be interpreted as confirming the absence of proportional bias, particularly given the limited sample size. The interrater differences between clinicians were small for the median per head (n=23; bias –0.020 to –0.036 cm; LoA~±0.5 to 0.6 cm), implying limited variability attributable to the clinician relative to the subgroup-dependent differences observed against the tape. However, these findings should be interpreted cautiously because only 2 clinicians participated. Both clinicians were blinded to the documented tape-based reference values and each other's measurements during the software-based measurements. Nevertheless, because EB had originally performed the tape-based measurements approximately 2 years earlier, recall of individual measurements cannot be completely excluded.

The 2 clinicians reported individual SUS scores of 92.5 and 95.0, which fall within the high range of established SUS benchmarks [18]. However, these results should be interpreted as preliminary descriptive feedback rather than as a formal usability evaluation. Further evaluation in a larger group of independent clinical users is needed, as the assessment was based on only 2 clinicians, both of whom were involved in the study and therefore were not fully independent end users.

Interpretation of Agreement With the Reference Measurement

Residual discrepancies between software-derived and tape-based HC measurements should be interpreted with caution, as tape measurements are operator-dependent (eg, landmark placement, tape tension, hair/scalp compression, and patient cooperation) and were not repeated. Tape-based HC is known to exhibit significant observer- and tape-related variability. Thakkar et al [8] note that a deviation greater than 0.5 cm in tape-based HC measurements is unacceptable from a clinical perspective. Yet they report that differences in observer- and repeat-measurements can exceed this value, even when using nonstretchable tapes. A small fraction of these differences reach 0.65 cm or greater intraobserver and 1.0 cm or greater interobserver. Because only one tape measurement was available per patient, the variability of the reference standard could not be estimated empirically or incorporated into the statistical analysis. Consequently, the observed software-versus-tape differences cannot be attributed exclusively to the software-based workflow and may partly reflect uncertainty in the tape-based reference measurement. This uncertainty should be considered when interpreting error metrics, such as MAE. For example, while the Delaunay MAE in nonirregular heads was 0.42 cm, its CI (0.28-0.58 cm) includes values above the 0.5 cm criterion. Occasional deviations above this threshold may therefore partly reflect real-world variability in the tape-based reference measurement. However, because repeated reference measurements were unavailable, reference-related and software-related measurement variability could not be separated empirically in this cohort.

Additionally, the cylindrical phantom experiment suggests that the implemented circumference calculation can reproduce the expected perimeter with high numerical precision under idealized software-based conditions. However, this experiment did not evaluate the full imaging-based workflow, including surface reconstruction from MRI data, nor did it assess the ability of Delaunay interpolation to mitigate surface irregularities. Therefore, it cannot determine how much of the observed discrepancy between software-derived and tape-based measurements arises from the tape-based reference versus the imaging-based workflow. The imaging-based method estimates a geometric perimeter derived from MRI intensities and may therefore be influenced by image quality, partial-volume effects at the skin boundary, motion,

and the selected iso-surface. The positive mean differences observed across subgroups indicate a small, systematic tendency toward larger HC values derived from the software in this cohort. This may reflect differences in the operational definitions of the 2 methods, namely an externally applied tape path with soft-tissue compression versus an image-derived outer-surface contour.

Comparison With Prior Work

Imaging-based approaches for retrospective HC estimation range from indirect geometric models and user-defined measurements on reformatted images to fully automated workflows embedded in clinical infrastructure. Vorperian et al [10] developed a model that adjusted imaging-derived cranial measurements for age-related growth patterns and differences between sexes. The model was derived from MRI examinations of 90 children and subsequently tested against contour tracings in 12 CT examinations. The model-based estimates did not differ significantly from the traced values. However, no directly comparable absolute error was reported; the validation sample was small, and no same-day tape measurements were available as a clinical reference.

Among the more automated approaches, Yepes-Calderón et al [12] describe an MRI-based pipeline integrated with PACS that extracts relevant masks, computes per-slice perimeter, selects the maximum HC, and automatically propagates the resulting value into the radiology report. In their evaluation, automated measurements differed from manual clinical HC by an average of 1.6 cm. Using a different modality but a similar axial-sampling method, Acamine et al [13] propose an automated CT-based approach that uses axial slices to detect the largest outer margin for HC measurement. This approach also extends to children with head abnormalities. They report a MAE of 1.26 cm between their CT-derived HC and head-tape measurements. In contrast to fully automated pipelines, Rau et al [11] report an MRI-based protocol using 3D sequences with multiplanar reformations, in which the axial plane is adjusted by the user to achieve the largest supra-auricular circumference, after which an ovoid region of interest is placed to approximate the outer head contour for circumference estimation. The method showed excellent interrater agreement, with an intraclass correlation coefficient of 0.985 and relative technical errors of measurement of 1.0% to 1.6%. Using a dedicated external surface scanner instead of radiological imaging, Beaumont et al [14] reported correlations between scanner-based and tape-based HC measurements, but both scanners consistently produced higher values, with maximum errors of 5% and 11%. Similar to this workflow, HC was derived from the intersection of a 3D head-surface model with a plane defined by 3 landmarks. However, their method required a dedicated external surface scan, used commercially available software (Rhinoceros 5, Robert McNeel & Associates), and did not separately assess variability related to manual landmark placement.

More recently, Adam et al [15] evaluated several MRI-based measurement strategies using the same cohort and reference measurements as this study. Their manual approaches and automated ellipsoid methods showed mean differences from the reference ranging from –0.524 to 0.097 cm, with SDs of 0.382 to 0.566 cm and r-TEM values of 0.55% to 1.12%. By contrast, the automated contour methods consistently overestimated HC, with mean differences of 2.365 to 2.486 cm, SDs of 0.414 to 0.491 cm, and r-TEM values of 3.42% to 4.21%. Seven neural network–based measurements were excluded because of gross segmentation failures. Klischat and Wilke [16] proposed an automated MRI-based approach using anatomical landmarks to estimate HC on obliquely oriented planes. In 153 pediatric MRI datasets, automated and manual measurements differed by a mean of 0.044 cm with a median absolute deviation of 0.8349 cm. However, the method may be less reliable in cases with atypical cranial morphology because it depends on landmark transfer and a convex-hull assumption, and its performance in pathological cranial conditions remains uncertain because these were not specifically evaluated.

Although these approaches enable image-based HC estimation, the degree of clinician control differs considerably. Fully automated workflows allow limited case-specific adjustment, whereas reformat- and slice-based methods may remain sensitive to head position, image orientation, and atypical cranial morphology.

The current workflow provides a clinician-guided alternative, in which the user remains involved in selecting the plane and reviewing the resulting contour. This may be especially useful in anatomically complex cases, where fixed-plane or fully automated measurements may be affected by local surface distortions. The optional Delaunay-based regularization was specifically intended to reduce the influence of such irregularities on the measured perimeter.

When interpreting the reported performance, direct numerical comparison remains limited by differences in cohorts, imaging modalities, measurement protocols, reference standards, and error metrics. The MAEs of 0.42 to 0.43 cm observed in this study for heads without surface irregularities and an MAE of 0.45 cm obtained with the Delaunay-based regularization in cases with cranial surface irregularities were numerically lower than the MAE of 1.26 cm reported by Acamine et al [13] and the average difference of 1.6 cm observed by Yepes-Calderón et al [12], although these values are based on different error measures and should therefore be interpreted cautiously. Other studies primarily reported significance tests, interrater agreement, correlations, repeatability, or bias and therefore cannot be directly compared with these MAE values [10,11,14,16]. Adam et al [15] provide the closest contextual comparison because the same cohort and reference measurements were used. Their manual contour approaches showed little systematic bias, with mean differences of 0.023-0.097 cm and SDs of 0.508-0.531 cm, whereas the automated contour methods consistently overestimated HC with mean differences of 2.365-2.486 cm and SDs of 0.414-0.491 cm. Together with the lower relative technical error of measurement (r-TEM) values reported for the manual approaches and automated ellipsoid methods than for the automated contour methods, the findings suggest that both the level of automation and the method used to derive the contour

may affect measurement performance. They further indicate that contour-based methods may benefit from a clinician-guided review and case-specific adjustment. Direct comparison nevertheless remains limited because mean difference, r-TEM, and MAE capture different aspects of measurement performance and are not directly comparable.

Limitations

The evaluation has several limitations. The overall cohort size was modest, and the subgroup of participants with cranial surface irregularities included only 5 participants. This limits the precision and generalizability of subgroup estimates, despite the use of cluster-bootstrap confidence intervals. The limited sample size also reduced the ability of the proportional-bias analysis to detect true trends. This limitation is particularly relevant to the cranial surface irregularity subgroup, which comprised only 5 heads (10 rater-level observations), and therefore the nonsignificant slopes should not be interpreted as evidence that proportional bias is absent. In addition, only 2 clinicians participated, so although we quantified interrater agreement, future work should evaluate the approach across raters with different levels of experience and across additional sites.

The study relied on a single tape-based HC measurement per patient, obtained by one clinician. Given the known interobserver variability of tape-based HC measurements, the use of a single reference value introduces uncertainty and may contribute to the observed disagreement between methods [6-8]. In contrast, each clinician repeated the software-based measurement 3 times to assess repeatability and characterize variability in manual landmark placement. This asymmetry between the reference and index measurements should be considered when interpreting the agreement results. Both clinicians were blinded to the documented tape-based reference values during the software-based measurements. As EB had originally performed the tape-based measurements approximately 2 years earlier, recall of individual measurements cannot be ruled out entirely, although the long interval likely reduced this possibility.

Finally, the current workflow remains manual because clinicians must select the 3 points that define the measurement plane. This manual selection increases measurement time and limits throughput compared to semiautomated approaches, which propose an initial 3-point configuration for the clinician to accept or adjust slightly.

There were also limitations related to validating the surface reconstruction and circumference calculations. Because the analysis was based on previously acquired imaging data and we did not have access to imaging phantom datasets that may have been acquired during the same period as the datasets we used, we were able to verify the circumference calculations only by comparing them with a software-based cylindrical surface at 2 radii, independent of the actual study measurements. This experiment did not validate surface reconstruction from MRI data or the ability of Delaunay interpolation to mitigate surface irregularities. Ideally, in the future, the full pipeline, from image acquisition through surface reconstruction and measurement, should be validated using actual imaging phantom data, and this should also include simulating defects.

Surface reconstruction relies on iso-surface extraction from raw MRI intensities with potential case-specific threshold adjustment (which was, however, not applied in our study). Because MRI intensity values are not inherently standardized across scanners or field strengths, the use of a fixed threshold represents a practical simplification. Scanner-specific effects were not analyzed separately because the cohort was small and markedly imbalanced across scanner models, including only one examination acquired on the Avanto system and one on the Skyra system. Accordingly, this study does not allow conclusions regarding scanner-specific differences in measurement performance, and this should be investigated in larger, appropriately balanced cohorts. Delaunay-based interpolation is an explicit contour-regularization method that smooths local surface indentations. Although it showed a more stable agreement in the small irregularity subgroup, further validation is required before it can be recommended as a default output or before standardized criteria for its use can be established.

Future Directions

Future work should aim to improve standardization while maintaining the workflow's transparency to the clinician-in-the-loop. Promising directions include semi-automated initialization of landmark placement and/or plane proposals, followed by clinician confirmation. Another direction is to develop a more robust and reproducible surface extraction method. To reduce reliance on subjective visual contour acceptance and to formalize decision rules for interpolation, the workflow could use an objective consistency check based on the circumferences of the intersection and Delaunay-interpolated contours. If the discrepancy between these values exceeds a predefined threshold, the software could prompt the user to recommend the interpolated contour.

Future evaluations should also use repeated tape-based and software-based measurements under comparable conditions and include multiple independent and blinded observers to achieve a more balanced assessment. This would allow researchers to distinguish more clearly between tape-related and software-related measurement variability.

Larger, multi-site validations and a broader spectrum of cranial morphologies are necessary to assess generalizability and refine subgroup-specific performance estimates. Finally, given the small, systematic, positive bias observed in this cohort, future work could assess calibration strategies, such as bias correction learned on an independent dataset, to determine if such adjustments improve agreement without compromising robustness in irregular cases.

Conclusions

This study supports the feasibility of a clinician-in-the-loop workflow for estimating pediatric HC from routinely acquired 3D MRI scans. The workflow involves surface reconstruction, landmark-defined plane selection, and contour-based circumference computation. Measurements were obtained

in all 23 examinations, with small interrater differences. In examinations without cranial surface irregularities, both contour variants showed similar agreement with tape-based HC, with mean differences of approximately 0.41 cm. In the 5 examinations with cranial irregularities, the Delaunay-interpolated contour remained closer to the tape-based measurements than the intersection contour; however, this subgroup finding should be considered preliminary. Differences between software- and tape-based measurements should be interpreted cautiously because only one operator-dependent tape measurement was available per patient.

Generative AI Use

Generative AI tools were used in a limited capacity during manuscript preparation solely to assist with the translation of short passages from German to English using DeepL (DeepL SE) and the refinement of selected text passages using Grammarly (Superhuman Platform Inc). All AI-assisted outputs were carefully reviewed, revised, and validated by the authors to ensure accuracy, clarity, and consistency with the intended scientific content. No generative AI tools were used for study design, data analysis, interpretation of results, or formulation of scientific conclusions.

Funding

The study was carried out as part of the project “Interoperable and eXplainable clinical decision support (iXplain_CDS)” funded by the German Federal Ministry of Education and Research (BMBF) (grant no. 01ZZ2021).

Data Availability

The code developed for this study is available from Zenodo (10.5281/zenodo.22830523). The measurements reported in this study were performed using build 2025.12.08-1338 and the library versions reported in the “Software environment” subsection. The version deposited in Zenodo (2026.08.04-1718) is a later release. The deposited version was tested with newer library versions, as reflected in the current requirements.txt; these dependency updates did not alter the measurement or calculation workflow used in the study. The software also contains a placeholder for a future automation feature, which is not yet implemented. This placeholder was removed from the manuscript figures to avoid confusion about the software's current functionality. The code changes between the study build and the deposited version were limited to nonfunctional corrections to comments. The MRI images remained at the Institute of Diagnostic and Interventional Neuroradiology at Hannover Medical School; access to these data is restricted due to patient confidentiality.

Authors’ Contributions

Conceptualization: EB, UVJ, DW
Data Curation: LD, UVJ, EB, JPA
Software: LD, UVJ, DW
Resources: EB, JPA, DW
Formal Analysis: LD, DW
Supervision: EB, DW
Validation: EB, DW
Visualization: LD
Investigation: LD, JPA, UVJ, EB, DW
Writing - original draft: LD, UVJ, EB, DW
Writing – review & editing: LD, JPA, UVJ, EB, DW
Methodology: LD, UVJ, EB, DW
Project Administration: EB, DW

Conflicts of Interest

None declared.

References

1. Harris SR. Measuring head circumference: update on infant microcephaly. Can Fam Physician. Aug 2015;61(8):680-684. [Medline: 26505062]
2. Woods CG, Parker A. Investigating microcephaly. Arch Dis Child. Sep 2013;98(9):707-713. [doi: 10.1136/archdischild-2012-302882] [Medline: 23814088]
3. Nellhaus G. Head circumference from birth to eighteen years. Pediatrics. Jan 1, 1968;41(1):106-114. [doi: 10.1542/peds.41.1.106]
4. Aboagye EAO, Commeree AN. Microcephaly. Pediatr Rev. Jul 1, 2025;46(7):412-414. [doi: 10.1542/pir.2024-006440] [Medline: 40588276]
5. Accogli A, Geraldo AF, Piccolo G, et al. Diagnostic approach to macrocephaly in children. Front Pediatr. 2021;9:794069. [doi: 10.3389/fped.2021.794069] [Medline: 35096710]
6. Bartram JL, Rigby AS, Baxter PS. The “Lasso-o” tape: stretchability and observer variability in head circumference measurement. Arch Dis Child. Aug 2005;90(8):820-821. [doi: 10.1136/adc.2004.063743] [Medline: 16040879]

7. Farkas LG. Accuracy of anthropometric measurements: past, present, and future. Cleft Palate Craniofac J. Jan 1996;33(1):10-18. [doi: 10.1597/1545-1569_1996_033_0010_aoampp_2.3.co_2] [Medline: 8849854]
8. Thakkar PA, Yagnik K, Parmar NT, Das RR, Thakkar UP. Observer variability in head circumference measurement using routine versus non-stretchable tapes in children. J Nepal Paedtr Soc. 2018;37(3):238-243. [doi: 10.3126/jnps.v37i3.19556]
9. Eichhorn H, Vascan AV, Nørgaard M, et al. Characterisation of children's head motion for magnetic resonance imaging with and without general anaesthesia. Front Radiol. 2021;1:789632. [doi: 10.3389/fradi.2021.789632] [Medline: 37492164]
10. Vorperian HK, Durtschi RB, Wang S, Chung MK, Ziegert AJ, Gentry LR. Estimating head circumference from pediatric imaging studies an improved method. Acad Radiol. Sep 2007;14(9):1102-1107. [doi: 10.1016/j.acra.2007.05.012] [Medline: 17707318]
11. Rau A, Demerath T, Kremers N, Eckenweiler M, von der Warth R, Urbach H. Measuring the head circumference on MRI in children: an interrater study. Clin Neuroradiol. Dec 2021;31(4):1021-1027. [doi: 10.1007/s00062-021-01019-z] [Medline: 34019101]
12. Yepes-Calderon F, Wihardja F, Sloan A, Kim J, Nelson MD, McComb JG. Measuring maximum head circumference within the picture archiving and communication system: a fully automatic approach. Front Pediatr. 2021;9:608122. [doi: 10.3389/fped.2021.608122] [Medline: 34350141]
13. Acamine PS, Maffei Loureiro R, dos Anjos Longas L, et al. An automated measurement of head circumference using CT scans: an application in children with head abnormalities. Neuroscience Informatics. Sep 2025;5(3):100217. [doi: 10.1016/j.neuri.2025.100217]
14. Beaumont CAA, Knoops PGM, Borghi A, et al. Three-dimensional surface scanners compared with standard anthropometric measurements for head shape. J Craniomaxillofac Surg. Jun 2017;45(6):921-927. [doi: 10.1016/j.jcms.2017.03.003] [Medline: 28454666]
15. Adam JP, Spineli LM, Winther HB, Bültmann E. Head circumference assessment in pediatric MRI: a pilot study of manual measurement methods and automated segmentation-based alternatives. Front Neurosci. 2026;20:1822095. [doi: 10.3389/fnins.2026.1822095] [Medline: 42368201]
16. Klischat J, Wilke M. An automated approach to extracting head and brain circumference from MRI datasets. PLoS One. 2026;21(7):e0352445. [doi: 10.1371/journal.pone.0352445] [Medline: 42507676]
17. Ramin L, Jan U, Wolff D. Head circumference tool (versions 2026.08.04-1718). Zenodo. 2026. URL: https://zenodo.org/records/22830523 [Accessed 2026-11-09]
18. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud. 2009;4(3):114-123. [doi: 10.5555/2835587.2835589]

Abbreviations

2D: two-dimensional
DICOM: Digital Imaging and Communications in Medicine
HC: head circumference
ITK: Insight Segmentation and Registration Toolkit
LoA: limits of agreement
MAE: mean absolute error
MPRAGE: magnetization-prepared rapid acquisition with gradient echo
MRI: magnetic resonance imaging
PACS: Picture Archiving and Communication System
r-TEM: relative technical error of measurement
SUS: System Usability Scale
VTK: Visualization Toolkit

Edited by Andrew Coristine; peer-reviewed by Hayk Stepanyan, Masab Mansoor; submitted 09.Apr.2026; initial comments to author 21.Jul.2026; final revised version received 16.Oct.2026; accepted 20.Oct.2026; published 20.Nov.2026

Please cite as:
Dittrich L, Adam JP, von Jan U, Bültmann E, Wolff D
Magnetic Resonance Imaging–Based Measurement of Pediatric Head Circumference by a Clinician-in-the-Loop 3D Surface Tool: Development and Validation Study
JMIR Med Inform 2026;14:e97692
URL: https://medinform.jmir.org/2026/1/e97692

doi: 10.2196/97692

© Leonie Dittrich, Jan Philipp Adam, Ute von Jan, Eva Bültmann, Dominik Wolff. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 20.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.