



Convolutional Neural Network for Fully Automated Cerebellar Volumetry in Children in Comparison to Manual Segmentation and Developmental Trajectory of Cerebellar Volumes

Daria Juliane Sobootian¹ · Paul Bronzlik¹ · Loukia M. Spineli² · Lena Sophie Becker³ · Hinrich Boy Winther³ · Eva Buelmann¹

Accepted: 25 September 2023
© The Author(s) 2023

Abstract

The purpose of this study was to develop a fully automated and reliable volumetry of the cerebellum of children during infancy and childhood using deep learning algorithms in comparison to manual segmentation. In addition, the clinical usefulness of measuring the cerebellar volume is shown. One hundred patients (0 to 16.3 years old) without infratentorial signal abnormalities on conventional MRI were retrospectively selected from our pool of pediatric MRI examinations. Based on a routinely acquired 3D T1-weighted magnetization prepared rapid gradient echo (MPRAGE) sequence, the cerebella were manually segmented using ITK-SNAP. The data set of all 100 cases was divided into four splits (four-fold cross-validation) to train the network (NN) to delineate the boundaries of the cerebellum. First, the accuracy of the newly created neural network was compared with the manual segmentation. Secondly, age-related volume changes were investigated. Our trained NN achieved an excellent Spearman correlation coefficient of 0.99, a Dice Coefficient of $95.0 \pm 2.1\%$, and an intersection over union (IoU) of $90.6 \pm 3.8\%$. Cerebellar volume increased continuously with age, showing an exponentially rapid growth within the first year of life. Using a convolutional neural network, it was possible to achieve reliable, fully automated cerebellar volume measurements in childhood and infancy, even when based on a relatively small cohort. In this preliminary study, age-dependent cerebellar volume changes could be acquired.

Keywords Cerebellum · Magnetic resonance imaging · Automated segmentation · Neural network · Pediatrics

Introduction

The cerebellum is well known to play an essential role in motor function. Because of its broad and complex involvement in cognition, emotional control, and social abilities, scientific interest in the cerebellum has grown, with multiple studies performed over the last decades [1–4]. Especially the developing cerebellum appears to be a challenging research

subject, considering the disturbance of its development may result in neurodevelopmental disorders and cognitive and neuromotor deficits [5, 6].

The human brain undergoes rapid maturation, especially in the first few years of life. The cerebellum is no exception. Commencing in utero in the early first trimester, cerebellar development requires roughly until the second postnatal year to achieve full circuit maturity [7]. The growth rate is awe-inspiring. In comparison to other brain structures, the cerebellum has the highest growth rate, more than doubling its size in the first 90 days after birth [8]. It continues to grow rapidly and thus reaches a total increase of 240% at the end of the first year of life, while the entire brain undergoes a percentage volume change of only 101% [9]. In the second year of life, growth flattens out, resulting in a total growth increase of 15% [9]. With the knowledge of the cerebellar volume changes during the first 2 years of life, this time is presumed to be highly vulnerable, highlighting the

Hinrich Boy Winther and Eva Buelmann contributed equally.

✉ Eva Buelmann
buelmann.eva@mh-hannover.de

- ¹ Institute of Diagnostic and Interventional Neuroradiology, Hannover Medical School, Hannover, Germany
- ² Midwifery Research and Education Unit, Hannover Medical School, Hannover, Germany
- ³ Institute of Diagnostic and Interventional Radiology, Hannover Medical School, Hannover, Germany

importance of pathology detection at an early stage in their development.

Due to its high spatial resolution and excellent image contrast without radiation exposure, MRI is the method of choice for diagnosing many clinical conditions in children. The cerebral development can be made visible [9], including physiological as well as pathological changes. Routinely, image analysis is performed qualitatively and is, therefore, highly dependent on the investigator. Quantitative analysis techniques such as T1 and T2 relaxation times, diffusion anisotropy, and magnetization transfer that support the interpretation of image data are desirable to detect discrete pathologies as early as possible and investigator-independent. Determining the cerebellar volume is an essential criterion in addition to describing its signal intensities and structure when assessing the cerebellum. Several studies have already addressed the relationship between cerebellar volume in different patient groups and various diseases or even certain environmental factors, such as ADHD [10–14]. Furthermore, the relations between the total volume of the brain and the volume of specific subdivisions, e.g., the cerebellum, in children with specific diseases have been repeatedly researched using different segmentation methods [15, 16]. These observations require precise volumetric measurements, which are very time-consuming, require a profound knowledge of cerebellar anatomy, and are also highly error-prone [17, 18]. Numerous methods for automated segmentation of medical imaging data have been introduced to address this challenging task over the years [17, 19, 20]. Multi-atlas label fusion techniques have emerged in the last decade. Even further development in that area was achieved with the introduction of deep learning techniques, such as the convolutional neural network presented in this study. Several studies applied these algorithms to calculate cerebellar volume [21, 22]. However, these studies were primarily performed in adults, given that segmenting cerebral structures in children, especially in newborns, is challenging due to their small size and reduced contrasts in imaging [23]. Furthermore, imaging data of inconspicuous healthy children are often not available.

The aim of the study was firstly to develop a fully automated and reliable volumetry of the cerebellum of children during infancy and childhood based on 3D T1-weighted images and deep learning algorithms and secondly to demonstrate the clinical significance of this investigator-independent, quantitative measurement.

Materials and Methods

The local institutional review board approved this retrospective study. All parents provided written informed consent to a scientific evaluation. From our database of pediatric

MRI examinations between 2014 and 2020, one hundred investigations scanned at 1.5T (Magnetom Aera, Siemens, Erlangen, Germany) using a 20-channel head coil were retrospectively selected, showing morphologically inconspicuous images without signal abnormalities as assessed by an experienced neuroradiologist (E.B., more than 15 years of experience). Brain MR imaging was performed for various clinical indications, including vomiting, headache, primary onset of epileptic seizures, tumor exclusion, or mild traumatic brain injury. Exclusion criteria were any signs of an infratentorial pathology and motion artifacts. Patient ages ranged between 0 and 16.3 years old (47 females, 53 males), with a mean age of 4 years and 50% younger than 24 months. Our routine MR protocol included a 3D T1-weighted magnetization prepared rapid gradient echo (MPRAGE) sequence with the following parameters: 1 mm continuous slice thickness, matrix 256×246 , TR 2200 ms, TE 2.67 ms, and flip angle 8° . The acquired data sets were converted into NIFTI files to use the open-source Segmentation program ITK-SNAP for postprocessing [24] (version 3.8.).

Segmentation Process

Manual segmentation was performed mainly in the axial plane. To have a standardized approach for delineating the brainstem from the cerebellum and to ensure that the cerebellar peduncles were correctly included, we used the method presented by Weier et al. [25]. Step one was to determine the image plane on which the cisternal part of the trigeminal nerve was visible. The same was done for the vestibulocochlear nerve. We then drew a line from each trigeminal nerve to the upper cerebellar peduncle (Fig. 1a) and in the extension of each vestibulocochlear nerve to the fourth ventricle (Fig. 1b), defining the rostral border of the cerebellum. Finally, the interpolation tool was used between these two layers to determine the rostral segmentation border in between (Fig. 1c).

Based on this delineation, the primary investigator, a specially trained graduate student, started each systematic manual segmentation of all 100 cases to teach the network the proper boundaries of the cerebellum. All 100 data sets were then manually segmented.

We utilized a four-fold cross-validation, dividing the 100 cases into four different splits. Each split contains 75 training and 25 testing cases. Each case was included once in the testing split, meaning no child was studied twice. The data selection for the four different splits was balanced by age. This was achieved by sorting the data sets by age and selecting each fourth entry using an offset of zero to three for the respective split.

Four separate neural networks were trained on each training split of the four-fold cross-validation. Two random exams of the respective training split were chosen to

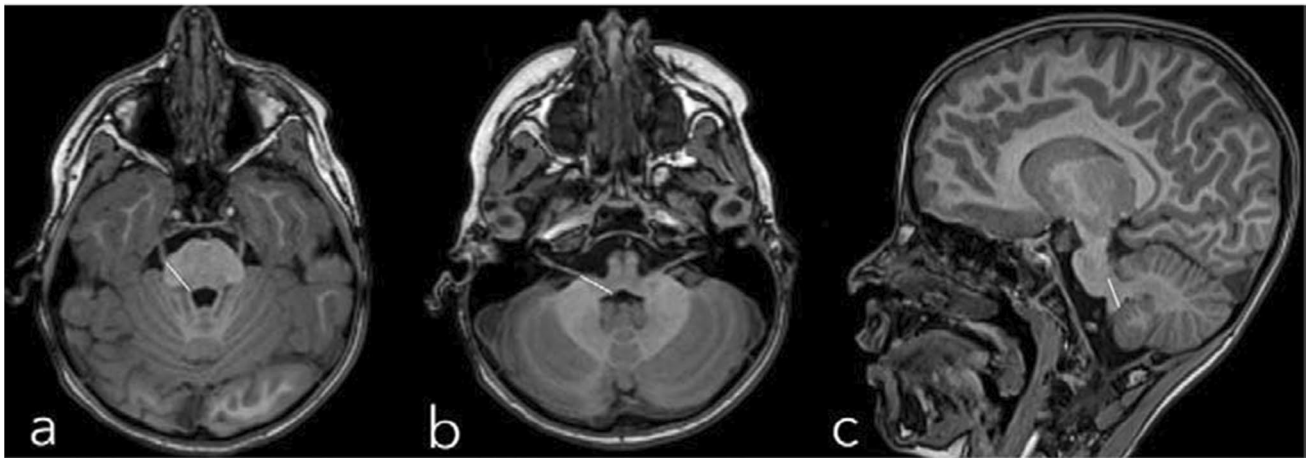


Fig. 1 a–c Standardized manual approach for delineating the brainstem from the cerebellum

calculate performance metrics during training. The best-performing model was selected as the final model for the respective split based on the training metrics.

Network Topology and Training

We used a 3D U-Net architecture derived from Çiçek et al. [26], as previously published [27]. It was implemented in TensorFlow [28]. The input layer has been expanded to $256 \times 256 \times 256$ neurons. Each downsample block consists of a batch normalization followed by two 3×3 convolutions with a 2×3 striding with zero padding. Each upsample block consists of a resampling layer, resizing the input to the same resolution as the corresponding downsample block using linear interpolation. This is followed by two 3×3 convolutions with zero padding. The output layer is the same size as the input layer. As suggested by He et al. [29], all activation functions have been converted from rectified linear units (ReLU) to parametric rectified linear units (PReLU). The neural network was trained by minimizing the weak binary cross-entropy as an objective function. We started with a learning rate of $1e-3$ and gradually lowered it to $1e-6$ during the training. Each epoch consisted of 200 virtual samples (steps) drawn dynamically from the image augmentation with a batch size of 2. We used linear transformations, such as rotation, shearing, and resizing, as well as non-linear transformations, such as local displacement for image augmentation.

The final model of each split was selected based on the performance metrics during training. The first model was trained for 113, the second for 159, the third for 386, and the fourth model for 143 epochs. We chose Adaptive Moment Estimation (Adam) as the stochastic optimization method. The training was performed on an eight-core Intel Xeon CPUE5-2637 v4 @ 3.50 GHz system with 128 GB

of RAM and two NVIDIA Tesla P40 GPUs with a total of 24 GB of video memory each and ran for 7 days.

Each of the resulting four networks was then used to perform an image segmentation of the 25 test cases for the respective split. No preprocessing, such as denoising or brain extraction, was performed.

To ensure reproducibility, 25 selected data sets (the testing split of one split of the four-fold cross-validation) were manually segmented a second time by the primary investigator. Here, too, a balanced age distribution was given. Furthermore, the same 25 data sets were assigned to an experienced neuroradiologist (E.B.) for manual segmentation to obtain inter-rater reliability.

Segmentation Quality Metrics

In this work, we provided metrics, such as the Dice coefficient, also known as Sørensen–Dice coefficient, and intersection over union (IoU), also known as the Jaccard index.

Volumetric Measurements

Based on the ground truth, the cerebellar volumes were calculated to obtain healthy children's age-dependent cerebellar volume growth. Unexpectedly large and small cerebellar volumes existed. We, therefore, examined their medical history in more detail and looked for reasons for these deviations. Based on their medical history, the outliers belonged to children with various neurodevelopmental disorders and concomitant micro- or macrocephaly, so we removed these 13 cases to describe normal age-related cerebellar development.

Statistical Analysis

We used violin plots with incorporated box plots to illustrate the volume distribution (in ml) for each sex and age group (age categorized into distinct, relatively balanced groups). We performed two multivariable ordinary least squares regression analyses: (i) regressing the volume against the age in years (in the logarithmic scale), sex, and head size (model 1), and (ii) removing head size from the model (model 2). We compared these two models regarding the adjusted R^2 and the Akaike information criterion (AIC). The model with the largest adjusted R^2 and a smaller AIC by 2 points was preferred. For the selected model, we performed regression diagnostics to inspect the appropriateness of the model assumptions. We reported the regression coefficients (measured in mean difference) and the 95% confidence interval: a confidence interval that excludes zero (the null value) coincides with a p -value less than 5% and indicates statistical significance. We created the scatterplot with the fitted line for each sex separately, using different point shapes for the head size. All analyses and visualizations were implemented using the statistical software R (version 4.2.1) [30]. We used the R-package *AICcmodavg* for the AIC results and the R-package *ggplot2* for the visualizations mentioned above [31, 32].

Comparison to Prior Work

We compared our results with the predictions of ACAPULCO and FreeSurfer. We use the CPU version 0.3.0 of ACAPULCO pediatric as a docker container provided by the original authors: registry.gitlab.com/shuohan/acapulco:0.3.0 (Image ID: a9a95b207bf8). FreeSurfer was utilized in version 7.4.1 as a docker container provided by the FreeSurfer project: freesurfer/freesurfer:7.4.1 (Image ID: 2ce88773a7f6).

Table 1 The Spearman rank-order correlation coefficient (based on the predicted and ground truth cerebellar volume), confidence intervals, Dice Coefficient, and IoU metrics have been calculated for the trained neural network (including the different splits of the four-fold

Variable	N	Est. (95% Conf. Int.)		Dice \pm SD	IoU \pm SD
Our Results					
Overall	100.0	0.99(0.98 to 0.99)	■	95.0 \pm 2.1 %	90.6 \pm 3.8 %
Split 1 (4-fold x-val)	25.0	0.98(0.96 to 0.99)	—■	95.5 \pm 2.3 %	91.4 \pm 4.1 %
Split 2 (4-fold x-val)	25.0	0.97(0.94 to 0.99)	—■	94.7 \pm 2.3 %	90.1 \pm 4.1 %
Split 3 (4-fold x-val)	25.0	1.00(0.99 to 1.00)	■	95.3 \pm 1.5 %	91.1 \pm 2.7 %
Split 4 (4-fold x-val)	25.0	0.99(0.97 to 0.99)	—■	94.5 \pm 2.3 %	89.7 \pm 4.1 %
Inter-rater	25.0	0.99(0.98 to 1.00)	—■	95.1 \pm 1.1 %	90.7 \pm 1.9 %
NN vs primary investigator	25.0	0.98(0.96 to 0.99)	—■	95.5 \pm 2.3 %	91.4 \pm 4.1 %
NN vs secondary investigator	25.0	0.98(0.95 to 0.99)	—■	94.0 \pm 2.2 %	88.7 \pm 3.8 %
Comparison to Prior Work					
ACAPULCO pediatric (v0.3.0)	100.0	0.88(0.83 to 0.92)	—■	68.1 \pm 37.0 %	61.3 \pm 35.2 %
FreeSurfer (v7.4.1)	86.0	0.93(0.89 to 0.95)	—■	85.5 \pm 16.8 %	77.4 \pm 18.5 %

Results

Automated Segmentation of the Cerebellum

As shown in Table 1, our trained neural network achieved an overall Spearman correlation coefficient of 0.99, a Dice Coefficient of $95.0 \pm 2.1\%$, and an IoU of $90.6 \pm 3.8\%$ with the ground truth as shown in Fig. 2. The mean absolute difference amounts to 2.4 ± 3.9 ml. The mean absolute percentage error (MAPE) is $3.1 \pm 3.2\%$.

The inter-rater correlation can also be derived from Table 1. We achieved a Spearman correlation coefficient of 0.99, a Dice Coefficient of $95.1 \pm 1.0\%$, and an IoU of $90.7 \pm 1.9\%$ from the 25 subjects examined by both our investigators. In addition, the accuracy of the newly created neural network was compared with the manual segmentation of our primary investigator and secondary investigator based on these 25 subjects. It showed a Spearman correlation coefficient of 0.98, a Dice Coefficient of $95.5 \pm 2.3\%$, and an IoU of $91.4 \pm 4.1\%$ for the neural network in comparison to the primary investigator and a Dice Coefficient of $94.0 \pm 2.2\%$, and an IoU of $88.7 \pm 3.8\%$ compared to the secondary investigator. These results are shown graphically in Fig. 2. Three examples of segmentations comparing the manual segmentation from investigator one and investigator two and the neural network's prediction are shown in Fig. 3.

Comparison to Prior Work

We compared our results to the pediatric parcellation protocol of the ACAPULCO pipeline [33] provided by Han et al. (version 0.3.0 CPU). ACAPULCO achieved a Spearman correlation coefficient of 0.883, a Dice coefficient of $68.1 \pm 37.0\%$ (mean \pm std), and an IoU of $61.3 \pm 35.2\%$ compared to the manual ground truth segmentation.

cross-validation), inter-rater-correlation, and comparison with the primary and secondary investigator as well as for ACAPULCO and FreeSurfer. The metrics of ACAPULCO and FreeSurfer are in regard to the manual ground truth segmentation

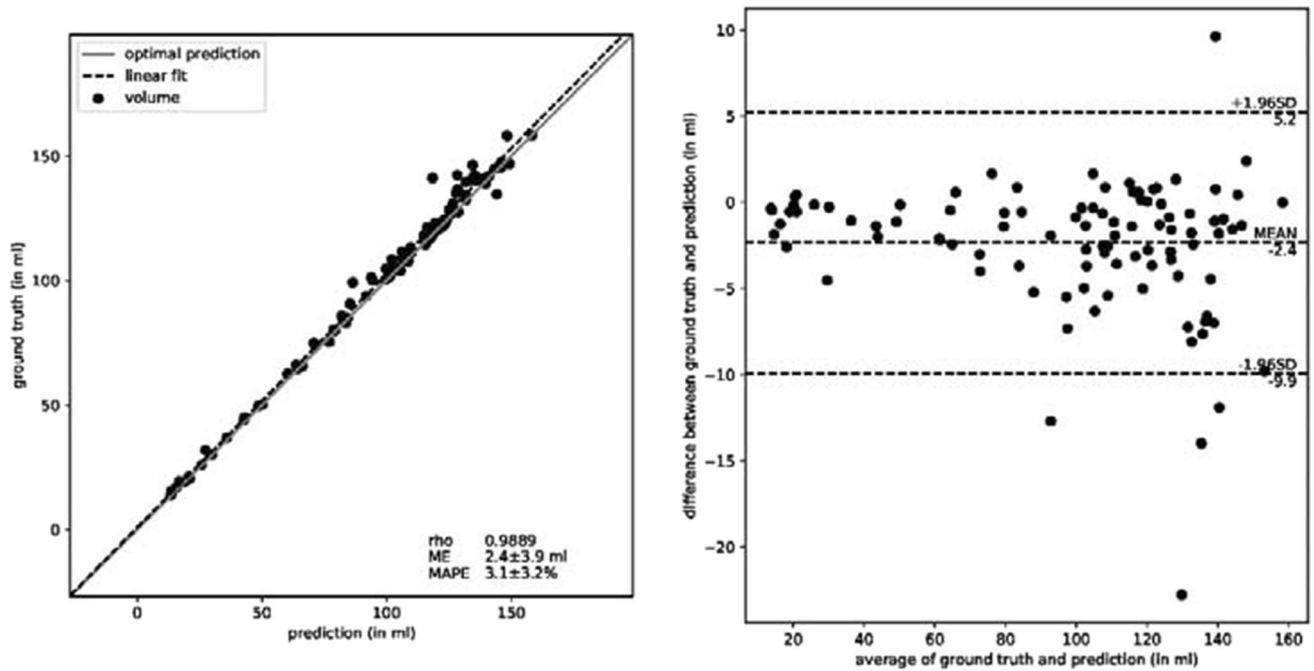


Fig. 2 Display of congruence between the ground truth volume, determined by the manual segmentation, and the volume prediction by the neural network

FreeSurfer reached a Spearman correlation coefficient of 0.93, a Dice Coefficient of $85.5 \pm 16.8\%$, and an IoU of $77.4 \pm 18.5\%$ compared to the manual ground truth segmentation. A total of 86 cases were processed successfully. Of the 14 missing cases, 12 failed the Talairach check, one triggered a watershed error due to a failed brain region detection, and the skull stripping failed for one case. The 14 missing cases were omitted when calculating the IoU and volume metrics. Twelve of the 14 missing cases were under 1 year old.

Age-Dependent Manually Segmented Volumetric Measurements

Table 2 describes the relationship between cerebellar volume and age, demonstrating continuous growth. It illustrates the exponentially rapid growth in the first year of life. For male children, the calculated mean total cerebellar volume is higher in each age group, except for infants aged 0 to 6 months.

The panel of violin plots also indicates a tendency for a larger volume in males than females. However, the violins with box plots overlap to some extent, and the pattern persists over the age groups (Fig. 4). The volume in both sexes increases exponentially with age, as expected.

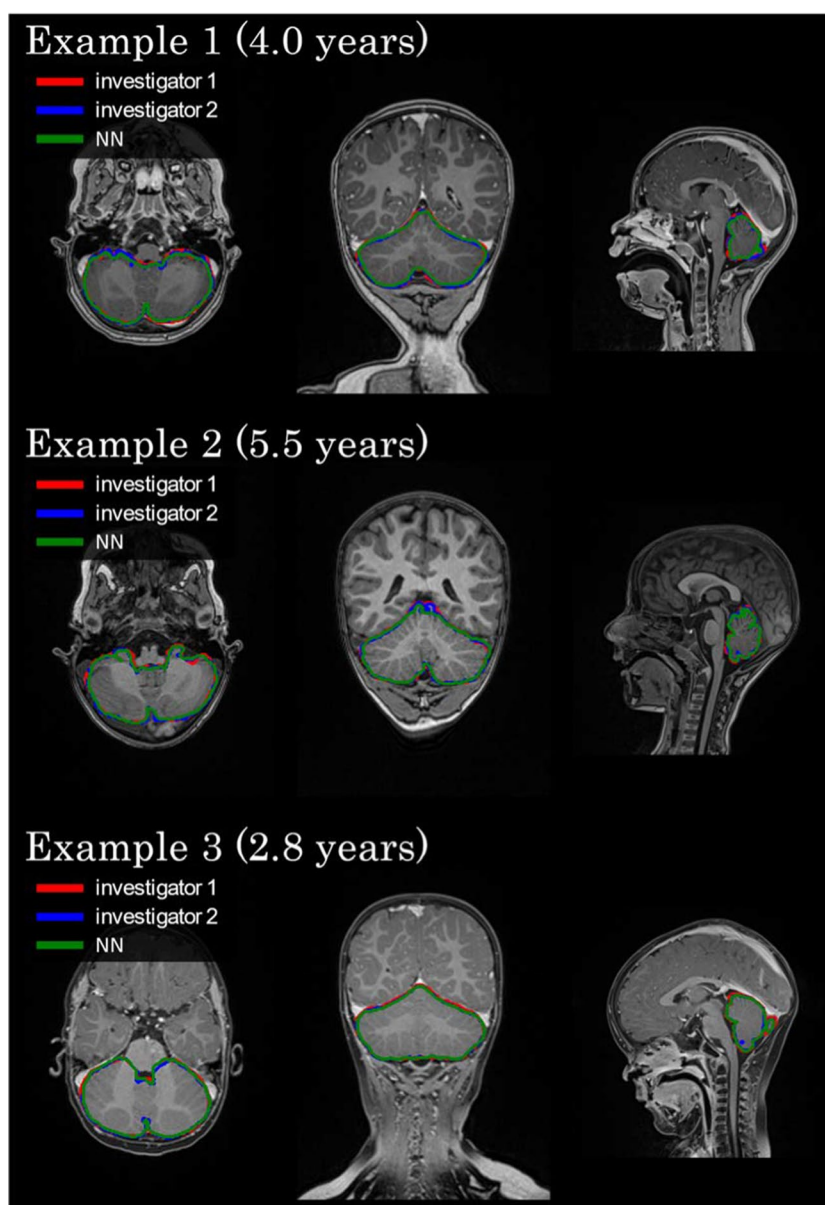
Regression Analysis

Model 1 was associated with a larger adjusted R^2 and a substantially smaller AIC value (by 31.18 points); hence, it was considered in the present study (Table 3).

The point estimate of the regression coefficients indicated a substantial association of the volume with all variables; however, there was a statistically significant association only with age and head size. Specifically, an older child had a larger volume on average than a younger child. Similarly, the volume increased by 29.46 ml on average (ranging from 17.26 to 41.65 ml) in a child with macrocephaly compared to normal head size. On the contrary, the volume decreased by 19.54 ml on average (ranging from 9.60 to 29.47 ml) in a child with microcephaly compared to normal head size. The results of the regression diagnostics can be found in the Supplementary material.

Figure 5 depicts the individual cerebellar volume of each examined child depending on the age at the time of the MRI. The fitted regression lines by sex illustrate the exponential growth of the volume as the child grows, increasing quickly during the first 2 years of life. Consecutively, the growth rate decreases sharply, rendering the curve much flatter. A distinct age dependency of the cerebellar volume can be inferred: with increasing age in infancy, the cerebellar volume increases considerably. The lines are inseparable at the

Fig. 3 Three examples of segmentations depicting the manual delineation from investigator one and investigator two and the neural network's prediction



earliest age but start being separated after the first 6 months of life, with males tending to have a systematically larger volume than females. However, there are a few children with very low volume specifications below the expected curve that attract attention.

Subgroup Analysis

The subgroup analysis, as depicted in Table 4, shows a statistically significantly worse prediction quality in the age group 0 – 6 months with a mean \pm std dice coefficient of 0.93 ± 0.02 , while the age groups 12 – 24 months and 2 – 4 years show a statistically significantly better prediction quality with a dice coefficient of 0.96 ± 0.02 and 0.96 ± 0.02 respectively. No statistically significant difference in quality

is observed in the sex subgroups or regarding the different splits of the four-fold cross-validation.

Discussion

One of the first approaches for a structured volume calculation was atlas-based, meaning the creation of different maps as a reference point for clinicians and investigators wanting to study the cerebellum [34]. While this segmentation tool for volume determination has been widely used for a long time, fully automated ones have only recently been explored [34–36]. A vast range of techniques and methods have already been applied, but up to date, the available automated methods for brain segmentation are mainly designed

Table 2 Mean cerebellar volume for the different age groups

Age groups	n (m:f)	Total cerebellar volume (ml)				
		Range	Mean	Mean male	Mean female	SD
Age-dependent mean cerebellar volume for all 100 patients						
0–6.0 months	23 (14:9)	13.90–74.81	35.97	34.71	37.94	19.80
6.1–12.0 months	11 (7:4)	64.63–126.72	93.84	99.32	84.25	17.28
12.1–24 months	16 (8:8)	74.20–128.48	106.67	111.78	101.56	14.39
2–4 years	20 (10:10)	82.83–158.30	121.23	126.46	116.01	16.05
4–6 years	10 (5:5)	111.18–146.39	130.72	138.34	123.11	11.95
6–16 years	20 (9:11)	79.93–158.11	132.66	137.39	128.79	17.77
Total	100					
Age-dependent cerebellar volume for normocephalic patients without outliers						
0–6.0 months	23 (14:9)	13.90–74.81	35.97	34.71	37.94	19.80
6.1–12.0 months	9 (5:4)	64.63–108.79	89.59	93.87	84.25	14.65
12.1–24 months	14 (8:6)	84.84–128.48	107.46	111.78	101.70	11.04
2–4 years	14 (6:8)	104.80–134.24	118.40	119.89	117.28	8.45
4–6 years	10 (5:5)	111.18–146.39	128.36	133.62	123.11	11.14
6–16 years	17 (8:9)	118.18–158.11	139.67	143.17	136.56	9.01
Total	87					

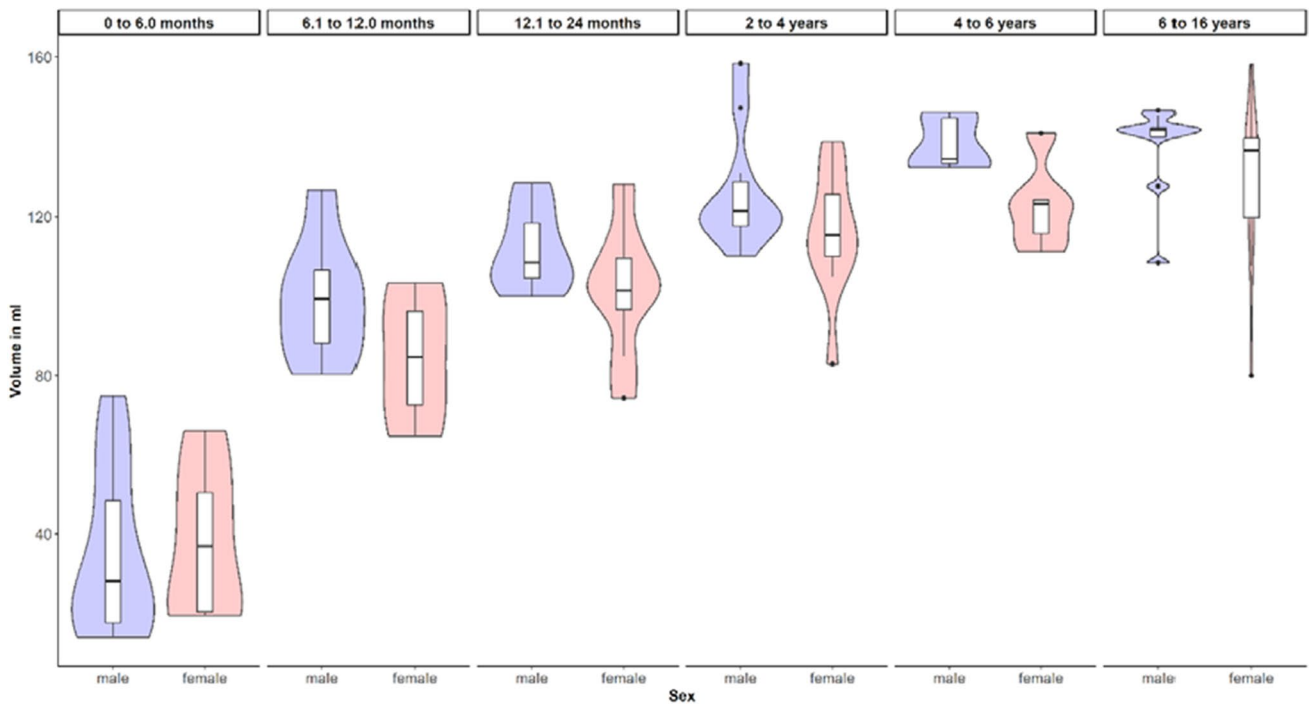


Fig. 4 The panel of violin plots indicated a tendency for a larger volume in males than in females

for adults instead of the pediatric population [37]. This is also reflected in our results when applying the FreeSurfer (version 7.4.1) to our data set, which yielded suboptimal results with a Dice coefficient of $85.5 \pm 16.7\%$ (mean \pm std) and a Spearman correlation coefficient of 0.93. While some studies, such as Tiemeier et al. [11], included children, determining the normal cerebellar volume during childhood

and adolescence, the parcellation was performed manually, and a distinctly higher age average was used compared to this study group with females being 13.7 years and males 13.9 years old. Narayanan et al. [38] introduced a probabilistic cerebellar atlas applicable to children’s data. However, the number of patients was limited to 18 subjects, and they included only children between 9 and 13 years.

Table 3 Results from both regression models

Variable	Model 1 ²	Model 2 ²
Log age ¹	<i>19.67 (18.28, 21.06)</i>	<i>19.22 (17.62, 20.84)</i>
Sex (female) ^{1,3}	<i>-5.19 (-10.54, 0.17)</i>	<i>-7.57 (-13.84, -1.31)</i>
Head size (microcephaly) ^{1,3}	<i>-19.54 (-29.47, -9.60)</i>	–
Head size (macrocephaly) ^{1,3}	<i>29.46 (17.26, 41.65)</i>	–
Adjusted R ²	90%	85%
AIC	808.45	839.63

AIC Akaike information criterion

¹Results refer to mean difference and 95% confidence interval (in parenthesis)

²Results in italics refer to statistically significant results (i.e., zero value of no difference is not included in the 95% confidence interval)

³Parenthesis indicates the non-reference level of the corresponding variable: “female” for sex, “microcephaly” and “macrocephaly” for head size

An additional development of the standard atlas-based segmentation is the multi-atlas label fusion. Open data from several atlases were used to improve accuracy [38]. For the neonatal brain, this has been presented by Otsuka et al. [39]. While a huge number of atlases is required to achieve accurate results with a broad age coverage, deep learning techniques, especially deep convolutional neural networks, have been demonstrated to produce meaningful results even with comparatively small data sets [20]. However, training on a limited data set increases the risk of overfitting the model on the training data and may hurt generalizability.

Different imaging techniques, as well as different organ systems, have already used neural networks [17, 40]. The applicability of neural networks for parcellation of the cerebellum has been explored by Carass et al. [41] and Han et al. [33] for the age group of approximately 8 to 13 years. However, to our knowledge, no neural network is applicable to infant cerebellar volumes of any age from birth to adolescence. This is also reflected in our results where we compared our results to the pediatric parcellation protocol of the ACAPULCO pipeline (Han et al. [33] version 0.3.0 CPU), which achieved poor results with a Spearman correlation coefficient of 0.883 and a Dice coefficient of $68.1 \pm 37.0\%$ (mean \pm std). The poor performance is most likely attributed to an overfit to the characteristics of the data set, i.e., ACAPULCO was trained on data from different MRI scanners (1.5T vs. 3T) with different protocol settings for the aforementioned age group.

The performance of our trained neural network is comparable—if not partly higher to similar studies with a Spearman correlation coefficient of 0.99 and a Dice coefficient of 95.0 [23, 38, 39, 42]. The results for the age-dependent subgroups were also excellent. The neonatal group (0–6 months old) performed statistically significantly worse compared to the other age groups, but even for these small babies with the lowest contrast due to incomplete myelination, the results were very good. Also, the calculated volumes of the

cerebella correspond to the values determined so far in the literature. Wu et al. [43] showed cerebellar development in children using volumetric calculations in different age groups. In the younger age categories, mean volume was calculated at 77.5 cm³ for the 3- to 11-month-olds and 104.7 cm³ for the 1- to 1.9-year-olds, very similar to the volumes determined by the present study (75.44 ml for 3 to 11 months and 106.99 ml for 1- to 1.9-year-olds). Differences are seen in the age group of 4- to 5.9-year-olds, where they present a mean volume of 113.6 cm³ in comparison to a mean volume of 128.36 ml in the present study. Furthermore, the volume calculated in our study continues to increase (6 to 8.9 years 136.87 ml and 9 to 12 years 142.43 ml), while Wu et al. showed no significant increase. This may be due to different case numbers, given that the present study included only four children aged 6 to 8.9 years and three between 9 and 12 years. Another possible reason could be the consideration of different ethnicities since only Chinese children were included in the aforementioned research. In contrast, we investigated mainly Caucasian children. Kosar et al. [44] also determined the cerebellar volume in healthy children and adolescents between 6 and 17 years old by using stereological volume measurements and applying the point counting technique. They measured an average total cerebellar volume of 123.44 cm³; our result of 139.67 ml is considerably higher. The strongly varying number of cases in these age groups can explain this discrepancy. Kosar et al. studied 90 children aged between 6 and 17 years. In contrast, the number of patients in the present study belonging to this age group is significantly smaller with 17 children (8 males and 9 females). In addition, the sex ratio of the included subjects may have influenced the lower volume. Of the 90 examined children by Kosar et al., 50 were female, known to have a smaller cerebellar volume [11].

The present study has limitations. Especially for children <2 years, the number of included cases is relatively small, and acquiring an even larger patient collective

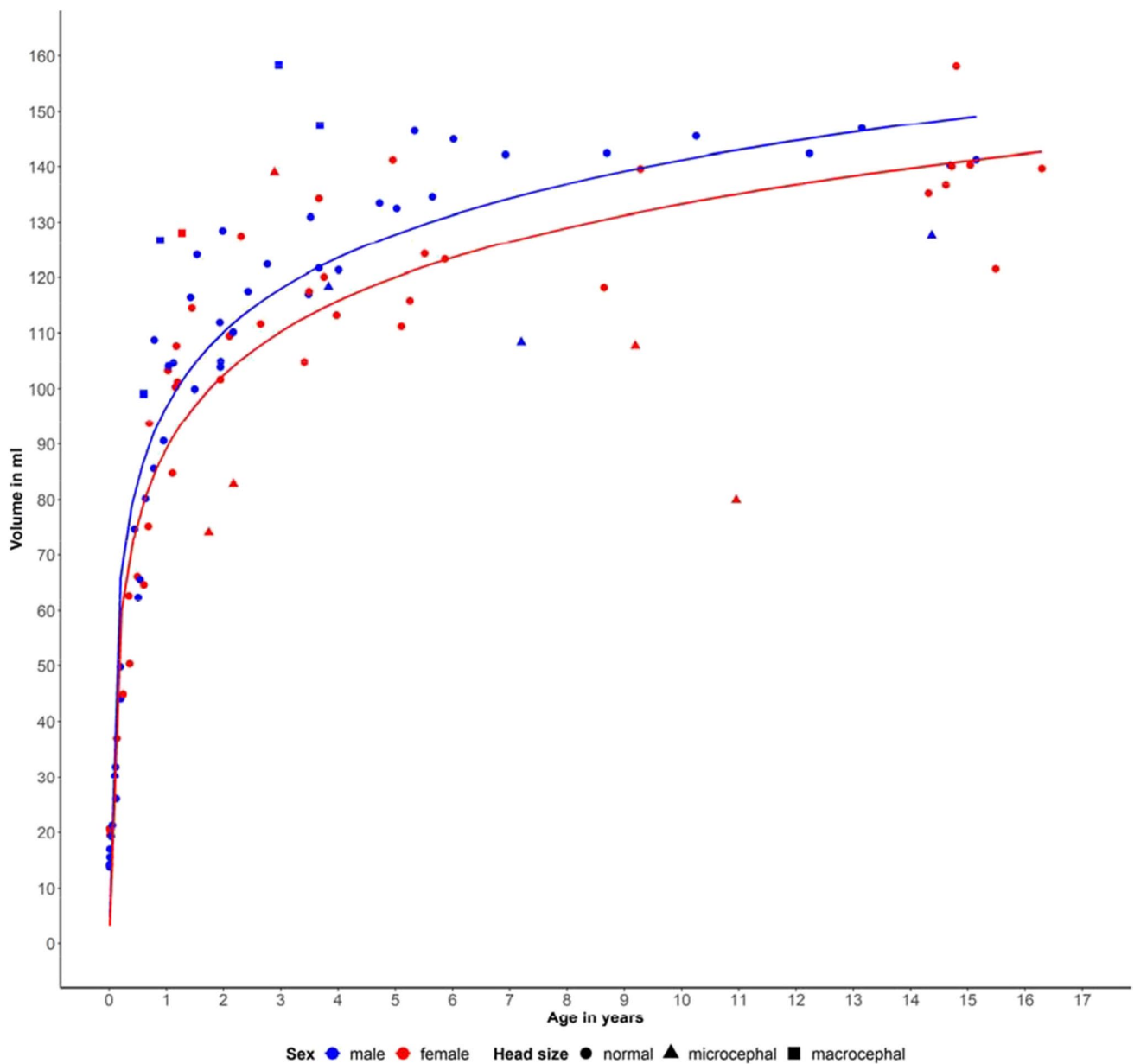


Fig. 5 Age-dependent cerebellar volume growth of female and male children between 0 and 16.3 years of age based on the ground truth

would be desirable. However, MRI data from morphologically inconspicuous children are difficult to achieve with known risks of prerequisite sedation or intubation. Furthermore, we included pediatric patients with various clinical diseases in order to obtain normal age-related cerebellar measurements. This makes the group inhomogeneous and can be erroneous due to a possible underlying neurological disorder that was not evident at the time of inclusion. Furthermore, segmenting infantile cerebella within the first year of life can be difficult due to the low contrast between different brain regions, which might complicate the clear demarcation from surrounding structures.

Additionally, we expect a certain degree of overfitting of the neural network, as this is a single-center study with a comparatively small and homogeneous data set.

Our goal was to generate a neural network that could reliably segment the cerebellar volume of children, which we could successfully prove despite the relatively small patient cohort. The calculated volumes corresponded well to those described in the literature. Furthermore, as shown by the outliers, quantitative analysis was superior to qualitative analysis because it could identify deviating volumes that were not visually recognizable.

Table 4 Subgroup analysis. Each subgroup was compared to the remaining data, e.g., split 1 was compared to split 2–4, respectively. The *p*-value was calculated using a Wilcoxon rank-sum test

Variable	N	Mean ± STD	P-value
Age			
0 - 6 Months	23	0.93 ± 0.02	2.6E-05 *
6 - 12 Months	11	0.95 ± 0.02	0.387
12 - 24 Months	16	0.96 ± 0.02	0.034 *
2 - 4 Years	20	0.96 ± 0.02	0.035 *
4 - 6 Years	10	0.95 ± 0.02	0.730
6 - 16 Years	20	0.95 ± 0.02	0.581
Sex			
Male	53	0.95 ± 0.02	0.666
Female	47	0.95 ± 0.02	0.666
4-Fold X-Val			
Split 1	25	0.95 ± 0.02	0.092
Split 2	25	0.95 ± 0.02	0.511
Split 3	25	0.95 ± 0.01	0.839
Split 4	25	0.95 ± 0.02	0.219

In summary, convolutional neural networks are a feasible technique to achieve reliable cerebellar volume measurements in childhood and infancy, even when based on a relatively small cohort. The volumes generated from clinical MRI investigations correlated well with manually acquired data and were comparable to published literature. In the future, using the developed neural network, cerebellar volumes can be measured in an acceptable time frame as part of routine MR imaging. Furthermore, if the presented method is applied to a larger group of healthy children, normal age-dependent references can be acquired, reflecting normal age-dependent cerebellar volume changes. Volumetric measurements based on this method could then act as an investigator-independent, quantitative biomarker of cerebellar development and help to improve the radiological diagnosis.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12311-023-01609-2>.

Author Contributions HBW and EB designed the study concept. Material preparation, data collection, and analysis were performed by DJS, PB, EB, and HBW. LS performed the statistical analysis. The first draft was written by DJS, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethical Approval and Consent to Participate We declare that this retrospective study was approved by the ethics committee of the Medical School of Hannover and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Hel-

sinki and its later amendments. For this type of retrospective study, formal consent is not required.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Pylayeva-Gupta Y. Consensus Paper: The cerebellum's role in movement and cognition. *Cerebellum*. 2015;23(1):151–77. <https://doi.org/10.1038/jid.2014.371>.
2. Schmahmann JD, Weilburg JB, Sherman JC. The neuropsychiatry of the cerebellum - insights from the clinic. *Cerebellum*. 2007;6(3):254–67. <https://doi.org/10.1080/14734220701490995>.
3. Gottwald B, Wilde B, Mihajlovic Z, Mehdorn HM. Evidence for distinct cognitive deficits after focal cerebellar lesions. *J. Neurol. Neurosurg. Psychiatry*. 2004;75(11):1524–31. <https://doi.org/10.1136/jnnp.2003.018093>.
4. Diamond A. Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Dev*. 2000;71(1):44–56. <https://doi.org/10.1111/1467-8624.00117>.
5. Riva D, Giorgi C. The cerebellum contributes to higher functions during development. Evidence from a series of children surgically treated for posterior fossa tumours. *Brain*. 2000;123(5):1051–61. <https://doi.org/10.1093/brain/123.5.1051>.
6. Allin M, Matsumoto H, Santhouse AM, Nosarti C, Alasady MHS, Stewart AL, Rifkin L, Murray RM. Cognitive and motor function

- and the size of the cerebellum in adolescents born very pre-term. *Brain*. 2001;124(1):60–6. <https://doi.org/10.1093/brain/124.1.60>.
7. Haldipur P, Dang D, Millen KJ. Embryology. *Handb. Clin. Neurol.* 2018;154:29–44. <https://doi.org/10.1016/B978-0-444-63956-1.00002-3>.
 8. Holland D, Chang L, Ernst TM, Curran M, Buchthal SD, Alicata D, Skranes J, Johansen H, Hernandez RY, Kuperman JM, Dale AM. Structural growth trajectories and rates of change in the first 3 months of infant brain development. *JAMA Neurol.* 2014;71(10):1266–74. <https://doi.org/10.1001/jamaneurol.2014.1638.Structural>.
 9. Knickmeyer RC, Gouttard S, Kang C, Evans D, Kathy Wilber J, Smith K, Hamer RM, Lin W, Gerig G, Gilmore JH. A structural MRI study of human brain development from birth to 2 years. *J. Neurosci.* 2008;28(47):12176–82. <https://doi.org/10.1523/JNEUROSCI.3479-08.2008>.
 10. Wyciszkievicz A, Pawlak MA, Krawiec K. Cerebellar volume in children with attention-deficit hyperactivity disorder (ADHD): replication study. *J. Child Neurol.* 2017;32(2):215–21. <https://doi.org/10.1177/0883073816678550>.
 11. Tiemeier H, Lenroot RK, Greenstein DK, Tran L, Pierson R, Giedd JN. Cerebellum development during childhood and adolescence: a longitudinal morphometric MRI study. *Neuroimage.* 2010;49(1):63–70. <https://doi.org/10.1016/j.neuroimage.2009.08.016>.
 12. Argyropoulou MI, Xydis V, Drougia A, Argyropoulou PI, Tzoufi M, Bassounas A, Andronikou S, Efremidis SC. MRI measurements of the pons and cerebellum in children born preterm; associations with the severity of periventricular leukomalacia and perinatal risk factors. *Neuroradiology.* 2003;45(10):730–4. <https://doi.org/10.1007/s00234-003-1067-0>.
 13. Dineen RA, Raschke F, McGlashan HL, Pszczolkowski S, Hack L, Cooper AD, Prasad M, Chow G, Whitehouse WP, Auer DP. Multiparametric cerebellar imaging and clinical phenotype in childhood ataxia telangiectasia. *Neuroimage Clin.* 2020;25(November 2019):102110. <https://doi.org/10.1016/j.nicl.2019.102110>.
 14. Weisenfeld NI, Peters JM, Tsai PT, Prabhu SP, Dies KA, Sahin M, Warfield SK. An MRI Study of Cerebellar Volume in Tuberous Sclerosis Complex. *Pediatr. Neurol.* 2013;48(2):105–10. <https://doi.org/10.1016/j.pediatrneurol.2012.10.011.An>.
 15. Juranek J, Dennis M, Cirino PT, El-Messidi L, Fletcher JM. The cerebellum in children with spina bifida and Chiari II malformation: quantitative volumetrics by region. *Cerebellum.* 2010;9(2):240–8. <https://doi.org/10.1007/s12311-010-0157-x>.
 16. Xavier Castellanos F, Lee PP, Sharp W, Jeffries NO, Greenstein DK, Clasen LS, Blumenthal JD, James RS, Ebens CL, Walter JM, Zijdenbos A, Evans AC, Giedd JN, Rapoport JL. Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. *J. Am. Med. Assoc.* 2002;288(14):1740–8. <https://doi.org/10.1001/jama.288.14.1740>.
 17. Grimm F, Edl F, Kerscher SR, Nieselt K, Gugel I, Schuhmann MU. Semantic segmentation of cerebrospinal fluid and brain volume with a convolutional neural network in pediatric hydrocephalus—transfer learning from existing algorithms. *Acta Neurochir.* 2020;162(10):2463–74. <https://doi.org/10.1007/s00701-020-04447-x>.
 18. Pierson R, Corson PW, Sears LL, Alicata D, Magnotta V, O’Leary D, Andreasen NC. Manual and semiautomated measurement of cerebellar subregions on MR images. *Neuroimage.* 2002;17(1):61–76. <https://doi.org/10.1006/nimg.2002.1207>.
 19. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. *J. Digit. Imaging.* 2019;32(4):582–96. <https://doi.org/10.1007/s10278-019-00227-x>.
 20. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med. Image Anal.* 2017;42(1995):60–88. <https://doi.org/10.1016/j.media.2017.07.005>.
 21. Cerasa A, Messina D, Nicoletti G, Novellino F, Lanza P, Condino F, Arabia G, Salsone M, Quattrone A. Cerebellar atrophy in essential tremor using an automated segmentation method. *Am. J. Neuroradiol.* 2009;30(6):1240–3. <https://doi.org/10.3174/ajnr.A1544>.
 22. Yang Z, Chuyang Ye JA, Bogovic AC, Jedynek BM, Ying H, Prince JL. Automated cerebellar lobule segmentation with application to cerebellar structural analysis in cerebellar disease. *Neuroimage.* 2016;127:435–44. <https://doi.org/10.1016/j.neuroimage.2015.09.032.Automated>.
 23. Anbeek P, Išgum I, Van Kooij BJM, Mol CP, Kersbergen KJ, Groenendaal F, Viergever MA, De Vries LS, Benders MJNL. Automatic segmentation of eight tissue classes in neonatal brain MRI. *PLoS One.* 2013;8(12):1–9. <https://doi.org/10.1371/journal.pone.0081895>.
 24. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* 2006;31(3):1116–28. <https://doi.org/10.1016/j.neuroimage.2006.01.015>.
 25. Weier K, Fonov V, Lavoie K, Doyon J, Louis Collins D. Rapid automatic segmentation of the human cerebellum and its lobules (RASCAL)—Implementation and application of the patch-based label-fusion technique with a template library to segment the human cerebellum. *Hum. Brain Mapp.* 2014;35(10):5026–39. <https://doi.org/10.1002/hbm.22529>.
 26. Çiçek Ö, Ahmed Abdulkadir S, Lienkamp S, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. *Medical Image Computing and Computer-Assisted Intervention.* 2016:424–32. <https://doi.org/10.1007/978-3-319-46723-8>.
 27. Winther HB, Gutberlet M, Hundt C, Kaireit TF, Alsady TM, Schmidt B, Wacker F, Sun Y, Dettmer S, Maschke SK, Hinrichs JB, Jambawalikar S, Prince MR, Graham Barr R, Vogel-Claussen J. Deep semantic lung segmentation for tracking potential pulmonary perfusion biomarkers in chronic obstructive pulmonary disease (COPD): The multi-ethnic study of atherosclerosis COPD study. *J. Magn. Reson. Imaging.* 2020;51(2):571–9. <https://doi.org/10.1002/jmri.26853>.
 28. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M. TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467 [cs]. 2016. <http://arxiv.org/abs/1603.04467>. Accessed 21 Jan 2017.
 29. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: surpassing human-level performance on imagenet classification. arXiv:1502.01852 [cs]. 2015. <http://arxiv.org/abs/1502.01852>. Accessed 2017-03-10.
 30. R Core Team (R Foundation for Statistical Computing). A language and environment for statistical computing. version 4.2.3. 2023. <https://www.r-project.org/>. Accessed 24 Aug 2023
 31. Mazerolle M. AICcmodavg: model selection and multimodel inference based on (Q)AIC(c). R package version 2.3-2. 2023. <https://cran.r-project.org/package=AICcmodavg>. Accessed 24 Aug 2023.
 32. Wickham H. ggplot2: Elegant Graphics for Data Analysis. US: Springer; 2016.
 33. Han S, Carass A, He Y, Prince JL. Automatic cerebellum anatomical parcellation using U-Net with locally constrained

- optimization. *Neuroimage*. 2020;218 <https://doi.org/10.1016/j.neuroimage.2020.116819>.
34. Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz A, Kabani N, Toga A, Evans A, Petrides M. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage*. 1999;10(3 1):233–60. <https://doi.org/10.1006/nimg.1999.0459>.
 35. Fonov V, Evans AC, Kelly Botteron RC, Almli CR, McKinstry, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*. 2011;54(1):313–27. <https://doi.org/10.1038/jid.2014.371>.
 36. Diedrichsen J. A spatially unbiased atlas template of the human cerebellum. *Neuroimage*. 2006;33(1):127–38. <https://doi.org/10.1016/j.neuroimage.2006.05.056>.
 37. Mayer KN, Latal B, Knirsch W, Scheer I, von Rhein M, Reich B, Bauer J, Gummel K, Roberts N, O’Gorman Tuura R. Comparison of automated brain volumetry methods with stereology in children aged 2 to 3 years. *Neuroradiology*. 2016;58(9):901–10. <https://doi.org/10.1007/s00234-016-1714-x>.
 38. Narayanan PL, Boonazier N, Warton C, Molteno CD, Joseph J, Jacobson JL, Jacobson SW, Zöllei L, Meintjes EM. Improved segmentation of cerebellar structures in children. *J. Neurosci. Methods*. 2016;262:1–13. <https://doi.org/10.1038/nbt.3301.Mammalian>.
 39. Otsuka Y, Chang L, Kawasaki Y, Dan W, Ceritoglu C, Oishi K, Ernst T, Miller M, Mori S, Oishi K. A multi-atlas label fusion tool for neonatal brain MRI parcellation and quantification. *J. Neuroimaging*. 2019;29(4):431–9. <https://doi.org/10.1111/jon.12623.A>.
 40. Minnema J, Wolff J, Koivisto J, Lucka F, Batenburg KJ, Forouzanfar T, van Eijnatten M. Comparison of convolutional neural network training strategies for cone-beam CT image segmentation. *Comput. Methods Prog. Biomed*. 2021;207 <https://doi.org/10.1016/j.cmpb.2021.106192>.
 41. Carass A, Cuzzocreo JL, Han S, Hernandez-castillo CR, Rasser PE, Ganz M, Beliveau V, Dolz J, Ayed IB, Desrosiers C, Thyreau B, Fonov VS, Louis Collins D, Ying SH, Onyike CU, Landman BA, Mostofsky SH, Thompson PM, Prince JL, et al. Comparing fully automated state-of-the-art cerebellum parcellation from magnetic resonance images. *Neuroimage*. 2018;150–72. <https://doi.org/10.1016/j.neuroimage.2018.08.003.Comparing>.
 42. Nie D, Wang L, Gao Y, Sken D. Fully convolutional networks for multi-modality isointense infant brain image segmentation. In: *Proceedings IEEE International Symposium on Biomedical Imaging*. IEEE Computer Society; 2016. p. 1342–5. <https://doi.org/10.1109/ISBI.2016.7493515>.
 43. Kuan-Hsun W, Chen C-Y, Shen E-Y. The cerebellar development in Chinese children—a study by voxel-based volume measurement of reconstructed 3D MRI scan. *Pediatr. Res*. 2011;69(1):80–3.
 44. Kosar MI, Karacan K, Otag I, Isleyen M, Gültürk S, Cimen M. Determination of cerebellar volume in children and adolescents with magnetic resonance images. *Folia Morphologica (Poland)*. 2012;71(2):65–70.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.