AJNR

Development of Gestational Age–Based Fetal Brain and Intracranial Volume Reference Norms Using Deep Learning

C.B.N. Tran, P. Nedelec, D.A. Weiss, J.D. Rudie, L. Kini, L.P. Sugrue, O.A. Glenn, C.P. Hess and A.M. Rauschecker

Development of Gestational Age–Based Fetal Brain and Intracranial Volume Reference Norms Using Deep Learning

C.B.N. Tran, P. Nedelec, D.A. Weiss, J.D. Rudie, L. Kini, L.P. Sugrue, O.A. Glenn, C.P. Hess, and A.M. Rauschecker

ABSTRACT

BACKGROUND AND PURPOSE: Fetal brain MR imaging interpretations are subjective and require subspecialty expertise. We aimed to develop a deep learning algorithm for automatically measuring intracranial and brain volumes of fetal brain MRIs across gestational ages.

MATERIALS AND METHODS: This retrospective study included 246 patients with singleton pregnancies at 19–38 weeks gestation. A 3D U-Net was trained to segment the intracranial contents of 2D fetal brain MRIs in the axial, coronal, and sagittal planes. An additional 3D U-Net was trained to segment the brain from the output of the first model. Models were tested on MRIs of 10 patients (28 planes) via Dice coefficients and volume comparison with manual reference segmentations. Trained U-Nets were applied to 200 additional MRIs to develop normative reference intracranial and brain volumes across gestational ages and then to 9 pathologic fetal brains.

RESULTS: Fetal intracranial and brain compartments were automatically segmented in a mean of 6.8 (SD, 1.2) seconds with median Dices score of 0.95 and 0.90, respectively (interquartile ranges, 0.91–0.96/0.89–0.91) on the test set. Correlation with manual volume measurements was high (Pearson $r = 0.996$, $P < .001$). Normative samples of intracranial and brain volumes across gestational ages were developed. Eight of 9 pathologic fetal intracranial volumes were automatically predicted to be >2 SDs from this age-specific reference mean. There were no effects of fetal sex, maternal diabetes, or maternal age on intracranial or brain volumes across gestational ages.

CONCLUSIONS: Deep learning techniques can quickly and accurately quantify intracranial and brain volumes on clinical fetal brain MRIs and identify abnormal volumes on the basis of a normative reference standard.

ABBREVIATIONS: GA = gestational age; IQR = interquartile range; SS-FSE = single-shot fast spin-echo

I n vivo fetal imaging, modalities such as ultrasonography and MR imaging, plays a central role in the assessment of fetal health and development during pregnancy. Fetal ultrasonography[1] and MR imaging have complementary strengths, with MR imaging

less limited by factors such as oligohydramnios, challenging fetal presentation, or acoustic shadowing from the ossifying calvaria.[2] MR imaging can also provide superior anatomic detail, which is an important consideration when assessing potential abnormalities of some fetal structures, especially in the fetal brain.[3] Thus, fetal brain MR imaging is performed in pregnant patients as early as 18 weeks of gestational age (GA), often after an anomaly is suspected on sonography, to provide further clarification for clinical management.[4] Fetal brain MR imaging is rapidly growing as a standard imaging technique for informing management decisions.

However, interpretation of fetal brain MR imaging remains a substantial challenge. Radiologic evaluation of fetal MR imaging is largely subjective and requires a high level of subspecialty expertise for consistent and accurate interpretation. Quantitative analysis is generally limited to 1D biometric measurements such as cerebral biparietal diameter or the transverse diameter of the
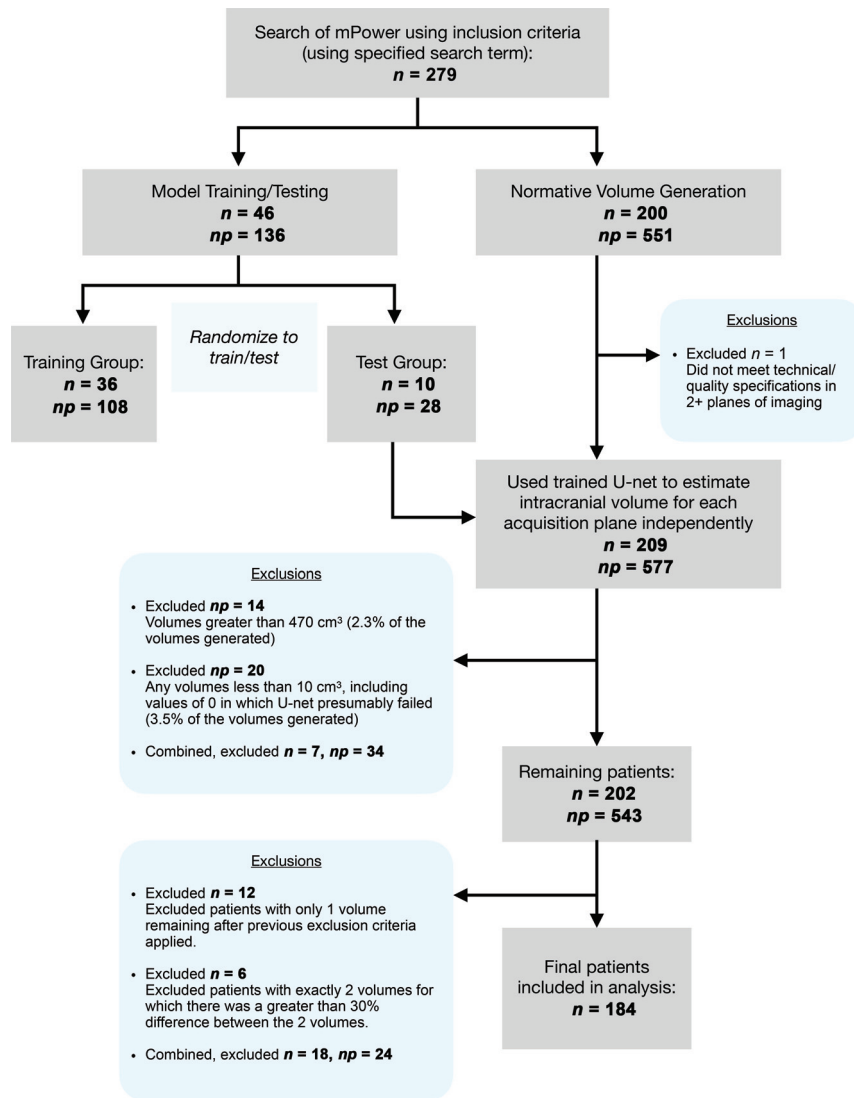
FIG 1. Flow chart shows study subject selection per exclusion criteria, from initial patient search to training set and test set randomization and development of a normative volume data set. *n* indicates the number of patients; *np*, number of planes of imaging.

would represent a major advancement in objective clinical and large-scale research assessments of fetal MR imaging.

We sought to develop a deep learning–based method for automated fetal brain MR imaging segmentation and volume quantification from single-shot fast spin-echo (SS-FSE) T2-weighted MRIs. Given the clinical relevance of fetal intracranial and brain volumes, analogous to routine 2D thecal and cerebral biparietal diameter measurements, our proof-of-concept study was specifically aimed at developing a method for fast and accurate measurement of these 3D volumes. Our ultimate goal was to apply the method to a large clinical population of fetal brain MRIs with normal findings to develop a normative reference for intracranial and brain volumes across a wide range of GAs.

MATERIALS AND METHODS
Definitions
Intracranial volume is defined as total volume within the cranium, including the brain, meninges, and CSF within the subarachnoid spaces and ventricles. Brain volume is defined as the combined volume of the brain parenchyma and ventricles. These 3D volume measurements are analogous to 2D thecal and cerebral biparietal/fronto-occipital diameters, which are currently used in clinical practice.

Patient and Data Selection
This retrospective study was approved by the institutional review board of the University of California, San Francisco, with a waiver of informed consent based on minimal risk. Included were a total of 246 study patients (ages, 16–45 years; median age, 33) who underwent fetal MRIs (SS-FSE T2WI) between 1999 and 2021.

Patients were identified by searching radiology reports in the institutional radiology archives (mPower; Nuance Communications). Inclusion criteria were patients with singleton pregnancies who underwent fetal brain MR imaging at our institution. Patients with abnormalities on their fetal MR imaging radiology report were excluded. Of the remaining patients, 46 patients were randomly sampled to constitute the training (*n* = 36) and test (*n* = 10) data sets (Fig 1, left column). The 10 patients included in the test sample were selected to include a representative range of GAs. The remaining 36 cases were assigned to the training sample, and their GAs were confirmed to also be representative of a wide range of GAs typical of the patients that undergo fetal MR imaging. For each patient, 1 optimal (non-motion-degraded) 3D

atria of the lateral ventricles.[5,6] Such manual measurements are performed as single linear measurements of 3D structures and are inherently subjective and prone to error.

While methods exist for quantitative evaluation of fetal MR imaging,[7,8] including a normative atlas of the fetal brain,[9] these methods typically use atlas-based segmentation techniques. Such techniques are well-suited for research questions but have limited clinical utility due to long processing times (hours) and frequent failure of segmentation in clinical cases that involve anatomic abnormalities.

Conversely, automated deep learning–based segmentation methods have the potential to consistently and objectively evaluate individual fetal intracranial and brain volumes in seconds. In particular, the U-Net architecture has proved highly effective for other biomedical segmentation tasks.[10,11] Obtaining quantitative estimates of intracranial/brain volume and deviations from normative data via automated deep learning methods

volume was manually chosen for each of the 3 planes of acquisition (axial, coronal, and sagittal).

An additional 9 patients' fetal MRIs were acquired using the search terms "microcephaly" and "macrocephaly" in the radiology report, to estimate fetal intracranial volume in cases of pathology in which intracranial volume was subjectively abnormal, as assessed by the pediatric neuroradiologist at the time of clinical interpretation.

Finally, an additional 200 patients' fetal brain MRIs were obtained through consecutive selection from the initial search, and these formed the data set for measuring normal intracranial volume (Fig 1, right column). This set of patients was selected to include only patients with normal intracranial and brain volumes for GA. However, to be representative of the clinical population, we allowed inclusion of patients with mild extracranial (eg, neck or spine) abnormalities ($n = 19$) and variations/mild abnormalities of intracranial structures (eg, mildly prominent subarachnoid or posterior fossa CSF spaces) ($n = 12$) but with normal biparietal measurements.

Additional Data from Chart Review

The GA of all fetuses at the time of MR imaging was obtained via chart review. In addition, maternal age at the time of MR imaging, the presence of maternal diabetes, and fetal sex were obtained via chart review when available, noting that not all auxiliary information was present for each patient.

MR Imaging Parameters and Ground Truth Segmentations

Ground truth intracranial and brain volume segmentations were based on manual segmentations performed by a medical student or radiology resident (C.B.N.T.) using ITK-SNAP (www.itksnap.org) and verified by a neuroradiologist (A.M.R., with 3 years' postresidency experience including with fetal brain MR imaging). Manual segmentations were independently performed in each plane of acquisition (axial, coronal, sagittal).

Imaging data were acquired using SS-FSE T2WI in a range of imaging parameters typical of clinical fetal brain MR imaging studies (TE range = 58–102 ms, median = 94 ms; TR range = 1–8 seconds, median = 4 seconds; in-plane resolution = 0.8–0.95 mm, median = 0.94 mm; section thickness = 2–4 mm, median = 3 mm).

Image Preprocessing

Images, while obtained using 2D acquisition protocols, were treated as 3D volumes by concatenating slices into a volumetric image. The advantage of treating images as 3D volumes is that information in slices adjacent to any particular 2D section may be informative for segmentation. These volumes were normalized by the mean signal intensity to zero mean and unit standard deviations (SDs). Individual acquisitions were resampled to a 1-mm^3 3D isotropic volume via linear interpolation. During training, elastic transformations[12] were applied to the images for data augmentation. These included small random rotations, translations, scaling, and free-form deformations. To fit within graphic memory constraints, the full-resolution augmented imaging volume was divided into 96 × 96 × 96 mm cubes (3D patches) as the network input. During training, the cubes were randomly sampled across the full-image volumes. The fetal intracranial contents may constitute only a relatively small portion of the entire MR image, which also includes portions of the mother's anatomy and other fetal anatomy. Therefore, to prevent sample imbalance, we sampled the same number of patches that included fetal intracranial voxels as those that excluded fetal intracranial voxels during training. A total of 60 patches were extracted from each training imaging volume ($n = 36 \times 3 = 108$), with 3 random augmentations per volume, resulting in 180 patches per volume or a total of 19,440 training patches. During testing, the MR imaging volume was densely sampled with the cubes using a step size of 32 mm in each direction, resulting in a 64-mm overlap between cubes. The overlapped segmentation predictions were averaged.

Convolutional Neural Network Model Architecture (U-Net) and Training

We used a 3D U-Net convolutional neural network architecture for segmentation of fetal brain MRIs. The same architecture was used unmodified from one previously developed to perform automated FLAIR lesion and intracranial metastases segmentations on MR imaging of the adult brain.[10,11] Our focus was on expanding the clinical application of this architecture by training it to perform intracranial and brain volume segmentations of fetal brain MR imaging. Thus, 2 sequential models were trained, the first for intracranial segmentation and a second for segmenting the brain from the intracranial volume identified by the first model. For the intracranial model, training was performed in 3 acquisition planes (axial, coronal, sagittal) when available across the 36 patients in the training set, treating each acquisition as an independent training sample, for 108 total training volumes. For the brain model, training was performed on a subset of 31 randomly chosen volumes using the same data augmentation parameters and a 6-fold cross-validation to create an ensemble model from a small training data set. For both models, we used a kernel size of 3 × 3 × 3, a dilation factor of 2 across all convolutional layers, and a batch size of twenty-four 3D patches. Cross-entropy loss and an Adam optimizer with a learning rate of 5×10^{-4} were used. The models were trained for 30 and 270 epochs, respectively. Hyperparameter optimization was not performed. Thresholding of the probability maps was set to 0.7 to decrease the false-positive rate relative to a threshold of 0.5. The network was implemented using TensorFlow 2 (www.tensorflow.org). Implementation was on a DGX-2 AI server, Version 4.5.0 (GNU/Linux 4.15.0–128-generic x86_64; NVIDIA).

Convolutional Neural Network Model Testing

Testing was conducted on 10 independent test patients, each with 2–3 acquisition planes, for a total of 28 test samples. During testing, the outputs of both models were postprocessed by taking the largest contiguous cluster of voxels and discarding the remaining predicted voxels to eliminate small false-positives outside the fetal intracranial contents. For the brain model, inner holes were also filled (scipy.ndimage.binary_fill_holes; SciPy 1.8.1; scipy.org).

The pretrained U-Nets were ultimately applied to a large set of fetal brain MRIs with clinically normal findings, as assessed by the radiology report, to develop a normative reference for intracranial and brain volume across GAs (Fig 2).
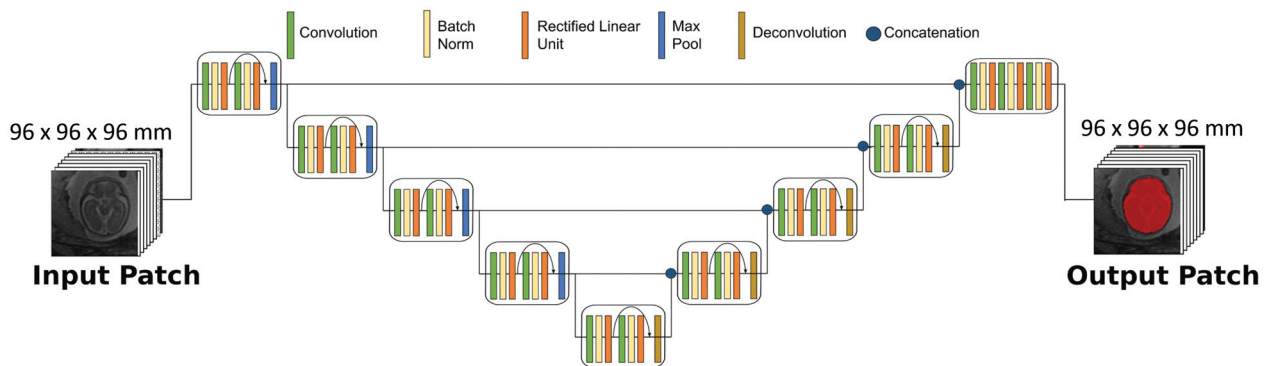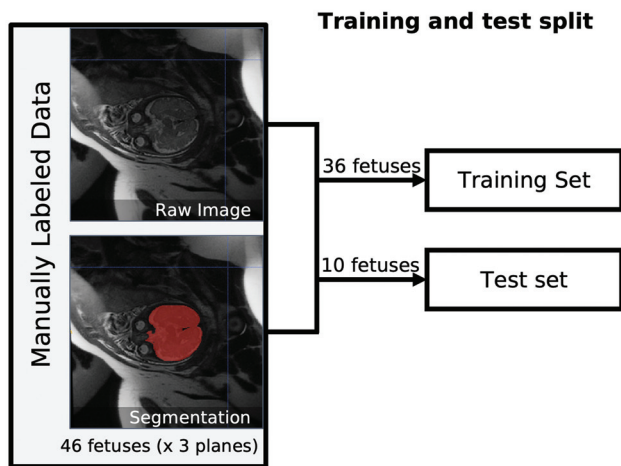
A   U-Net architecture
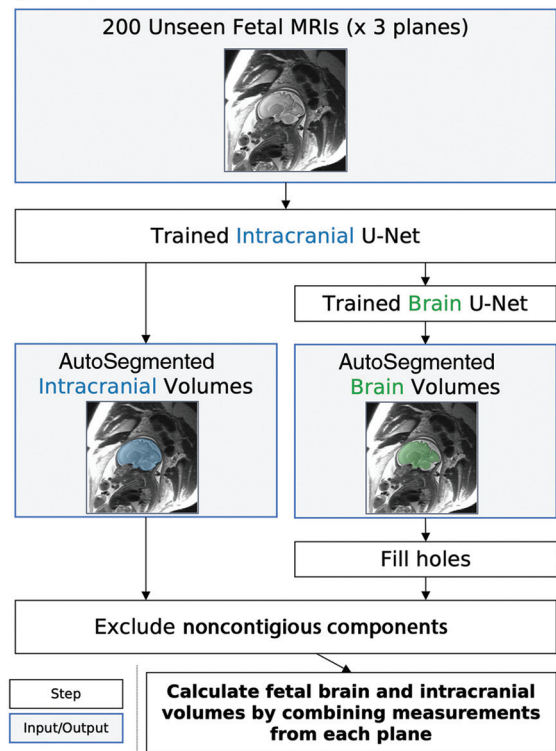
FIG 2. Method for training and testing the U-Nets. *A*, Schematic of 3D U-Net architecture used for training with sample input and output patch is shown. *B*, Manually segmented images were split into training (*n* = 36 patients) and test (*n* = 10 patients) sets. *C*, After confirmation of adequate segmentation performance on the test set, the trained U-Nets were applied to an additional 200 fetal brain MRIs for calculating normal fetal intracranial and brain volumes across GAs.

Normative Volume Generation

For the additional set of 200 fetal MRIs with normal findings without manual segmentation, a single high-quality image volume for each acquisition plane was chosen manually, as above. This process is very fast, taking approximately 3 seconds per volume. The volumes generated by the U-Net for the 10 patients in the test group were also included in the normative sample. For the brain model, an ensemble model was built by averaging all probability prediction maps of all 6 folds and then thresholding them at 0.7 to create the binary segmentation. Exclusion criteria (Fig 1) were applied separately to patients and acquisitions,

including patient MRIs that did not meet technical/quality specifications in at least 2 planes of imaging (eg, highly motion-degraded images or early termination of the examination) and generated volumes for each plane of imaging that were not biologically plausible (<10 cm^3 or >2 SDs of the entire distribution of volumes). Patients were then excluded from the normative sample if only a single generated volume remained or if exactly 2 volumes remained but there was a large ($>30\%$) discrepancy between the calculated volumes, thereby limiting confidence in the estimate. This process resulted in a final total of 184 patients in the normative reference group.

Volume Calculations

Predictions of intracranial and brain volumes were calculated from predicted segmentations independently for each imaging plane. To arrive at a single intracranial and brain volume estimate for each fetus, we combined these 3 data points for each patient applied independently to intracranial and brain volumes as follows: After applying the exclusion criteria, notably excluding biologically implausible statistical outlier volumes defined as <10 cm^3 or >470 cm^3 of intracranial volume or 373 cm^3 of brain volume (>2 SDs of entire underlying distribution), the median of the remaining 2–3 data points was used to estimate the total volume for each fetus. If only a single data point remained after exclusion criteria were applied, then this patient was excluded from the set of normative volumes (Fig 1).

Code Sharing

All code used for training and testing the model in this article has been made available at https://github.com/rauschecker-sugrue-labs/fetal-brain-segmentation. Code includes image preprocessing, model training, and model inference, including combining individual measurements from varying planes of acquisition into a final volume estimate.

Statistical Analysis

Segmentation Assessment. Performance of the intracranial U-Net was evaluated on the 28 test samples via comparison with manual segmentations. Performance of the brain U-Net was evaluated for each of the 6 folds (5–6 volumes each) on the data unseen by that fold. Dice scores were calculated for each test sample, comparing segmentations predicted by U-Net with the human reference standard. Mean and median Dice scores were assessed for each acquisition plane to assess the quality of automated segmentation by the acquisition plane. A Pearson correlation test was used to determine whether there was an association between GA and the Dice score.

Volumetry Assessment and Statistical Analysis. Total fetal intracranial and brain volumes calculated from processed U-Net outputs on their respective test samples were compared with those calculated from manual segmentations on the same test samples. A Pearson correlation test was used to determine the association between the volumes calculated from both methods. The root median square percentage error of those volumes was calculated, and 95% limits of agreement for each comparison were calculated. In addition, we performed a repeated measures ANOVA to assess whether any consistent pattern of differences in the volume calculations existed, depending on the acquisition plane chosen.

To investigate the effects of acquired demographic variables (fetal sex, maternal diabetes, and maternal age, binarized as $\geq$ or <35 years), we computed linear regressions for each demographic variable and for each category. To determine whether any significant differences existed between categories within a demographic variable, a bootstrapping technique was used. Specifically, the differences in the slopes and intercepts of best fit lines for 50,000 random permutations of these categories were computed, and a z

score and 2-tailed P value of the original data from this analysis are reported.

Stata/MP (Version 16; https://www.stata.com/features/overview/statamp/), Excel (Version 16.37; Microsoft), custom Matlab (R2019a; MathWorks) scripts, and custom Python (Version 3.8) scripts were used for these analyses. P values $< .05$ were considered statistically significant.

RESULTS

Patient Demographics

Training Set. GAs of the 36 fetuses comprising the training set ranged from 20.6 to 36.9 weeks (median, 24.6 weeks). Maternal age ranged from 19 to 40 years (median, 33 years). Nine (25%) fetuses were female, 19 (53%) were male, and 8 (22%) were of unknown sex. Four (11%) fetuses had a mother with a known history of diabetes.

Test Set. GAs of the 10 fetuses comprising the test set ranged from 20.7 to 36.1 weeks (median, 27.4 weeks). Maternal age ranged from 16 to 38 years (median, 33.5 years). Three (30%) fetuses were female, 5 (50%) were male, and 2 (20%) were of unknown sex. One (10%) fetus had a mother with a known history of diabetes.

Normative Values Set. GAs of the 184 fetuses comprising the final normative values set ranged from 19.9 to 37.7 weeks (median, 24.6 weeks). Maternal age ranged from 16 to 45 years (median, 33 years). Forty-eight (26%) subjects were female, 67 (36%) were male (including one XXY fetus), and 69 (37%) were of unknown sex. Twenty-four (13%) fetuses had a mother with a known history of diabetes.

Segmentation and Volumetry Performance on the Test Set. The convolutional neural networks resulted in highly accurate segmentations of fetal brains across a wide range of GAs (Fig 3). On test samples, overlap between segmentations predicted by the U-Nets and by human reference standards were near-perfect for both intracranial and brain segmentations. For intracranial segmentations, the median Dice score was 0.95 (interquartile range [IQR], 0.94–0.96) in the axial plane, 0.94 (IQR, 0.91–0.96) in the coronal plane, and 0.94 (IQR, 0.90–0.96) in the sagittal plane; and for brain, 0.90 (IQR, 0.89–0.90) in the axial plane, 0.89 (IQR, 0.88–0.90) in the coronal plane, and 0.90 (IQR, 0.90–0.91) in the sagittal plane (Fig 3B). A Pearson correlation test revealed a statistically significant positive association of $r = 0.49$ between GA and the Dice score ($P = .006$) for intracranial segmentations. A repeated measures ANOVA demonstrated that there was no consistent pattern of differences in the volume calculations, depending on the acquisition plane chosen ($P = .56$). There was no consistent bias of volumes generated by the automated-versus-manual intracranial segmentations (Fig 3C).

The automated method generated total intracranial volumes highly correlated with manual measurements (Pearson $r = 0.996$, $P < .001$). The volumes generated had low error, with a root median squared percentage error of 3.3% (Fig 3D). As expected, GA and estimated intracranial volume were correlated (Spearman $\rho = 0.92$, $P < .001$).
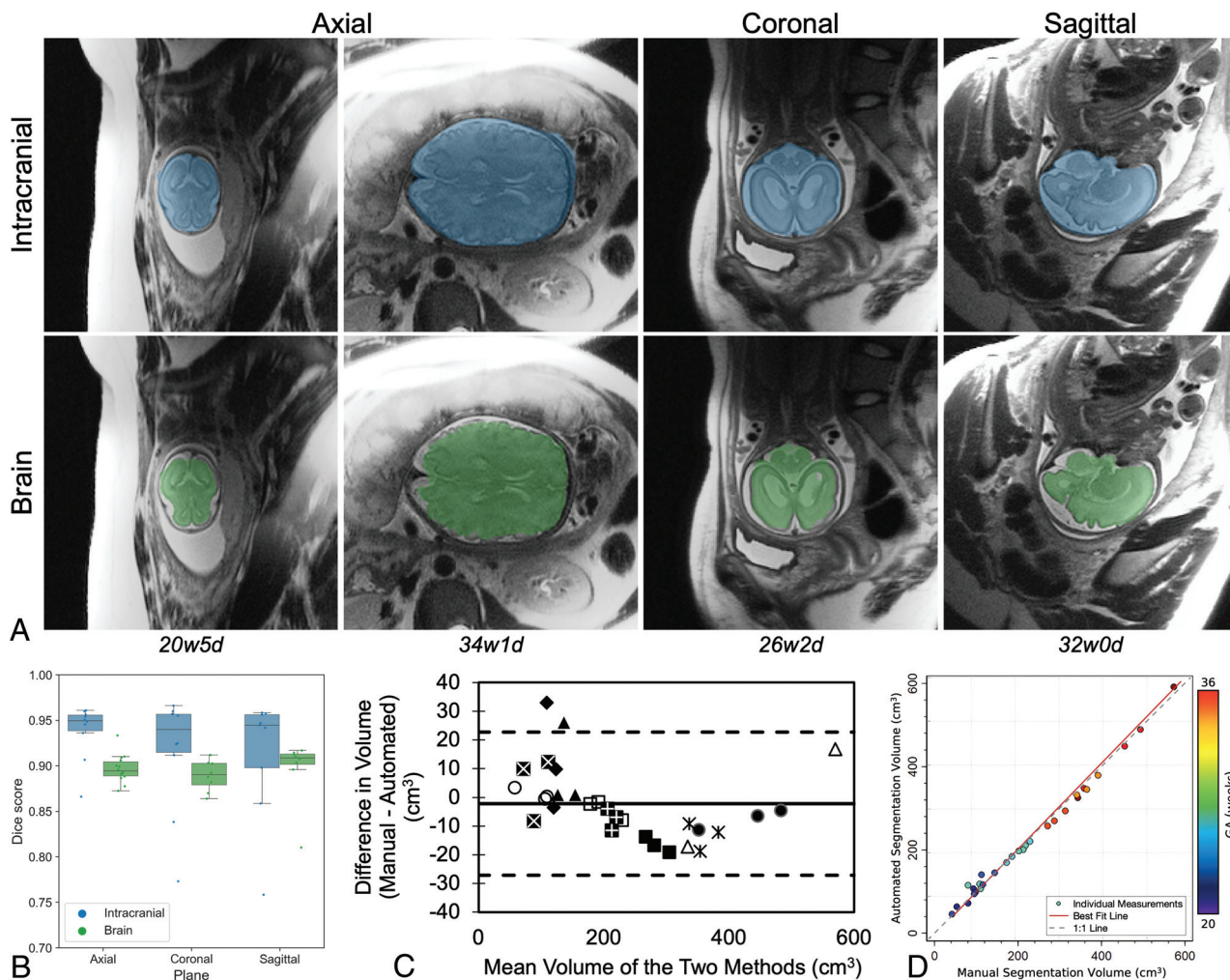
FIG 3. Performance of the U-Net for automated segmentation of intracranial (blue) and brain (green) volumes on the test set. *A*, Representative examples of the segmentation overlay on a section of the original brain MR imaging in various acquisition planes across multiple gestational ages (w = weeks, d = days). *B*, Individual Dice scores and boxplots compare the automated with the ground truth manual segmentation within axial, coronal, and sagittal dimensions, distinguishing scores for intracranial and brain segmentations. *C*, Bland-Altman plot demonstrates no linear trend in the difference between manual and automatically calculated intracranial volumes across the range of volumes tested. Each type of marker corresponds to axial, sagittal, and coronal measurements on an individual fetal brain. *D*, Scatterplot demonstrates strong agreement between manual and automated intracranial segmentation volumes, color-coded by GA. The best fit line and the 1:1 identity line are shown, nearly overlapping.

Timing of the U-Net

For an individual fetal brain MRI using our hardware, the U-Net produced an intracranial segmentation and associated mean volume in 6.8 (SD, 1.2) seconds on average. By comparison, an experienced human manually segmenting these volumes requires approximately 15 minutes for an accurate segmentation in an individual acquisition plane.

Normative Intracranial and Brain Volumes across GAs

To determine normal intracranial and brain volumes across a large population of fetal brain MRIs, we applied the trained 3D U-Nets to 209 clinically normal fetal brain MRIs ($n = 184$ after the exclusions described in Materials and Methods). As described in further detail in Materials and Methods, the models were applied in individual planes of acquisition (axial, sagittal, and coronal), which were then combined into 1 estimate of intracranial volume per fetus.

As expected, intracranial and brain volumes increased with GA, as did the variability of volumes (Fig 4 and Table). For example, the mean intracranial volume at 22–24 weeks was 103 (SD, 14) cm^3 ($n = 60$), while the mean intracranial volume at 32–34 weeks was 319 (SD, 48) cm^3 ($n = 17$).

To demonstrate the utility of this normative sample of automated fetal brain MR imaging volume measurements, we applied the trained intracranial U-Net to 9 pathologic brain MRIs (Fig 5A). Eight of the 9 volume calculations fell outside 2 SDs below or above the normative sample volumes (for GA), and the ninth volume was nearly 2 SDs above the average. We further investigated the relationship between GA and intracranial volume as a function of several demographic variables, such as fetal sex (Fig 5B), maternal age (Fig 5C), and the presence of maternal diabetes (Fig 5D). There were no appreciable effects of these demographic variables on the relationship between GA and intracranial volume in our data set, as shown by analyzing
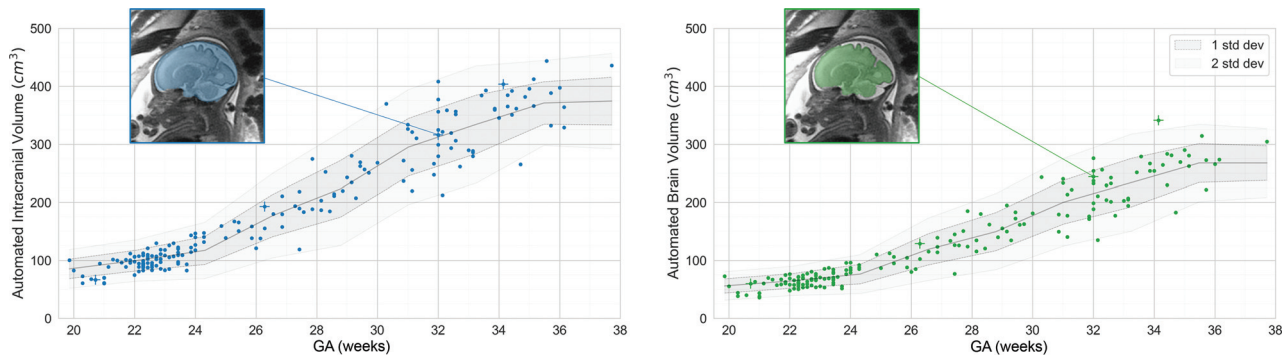
FIG 4. Intracranial (blue) and brain (green) volumes as a function of GA across 184 fetal MRIs with normal findings. Individual points represent automated measurements of volume in individual fetal brain MRIs. The *center line* represents a moving average across these points, ±1 (*dark shading*) or 2 (*light shading*) SDs. One fetal brain is shown as an example in the insets. Images shown in Figure 3A are denoted by a +.

Intracranial (n = 184) and brain (n = 178) volumes (5th, 50th, and 95th percentiles) as a function of GA across a set MRIs of fetuses with clinically normal brains, grouped by 2-week intervals

GA (Weeks)	Intracranial Volume			Brain Volume		
	5th	50th	95th	5th	50th	95th
19.5–21.5	60.4	71.2	100.8	37.1	49.7	70.8
21.5–23.5	82.8	99.8	122.2	53.1	65.5	80.5
23.5–25.5	91.8	139.5	166.0	54.7	86.9	110.6
25.5–27.5	120.2	169.1	213.5	78.8	114.6	141.2
27.5–29.5	184.4	217.0	276.0	119.4	150.4	187.5
29.5–31.5	229.1	269.9	350.1	145.2	180.2	241.7
31.5–33.5	248.9	306.0	383.4	172.0	206.8	255.8
33.5–35.5	321.4	366.3	406.3	213.2	268.1	307.9

the differences in best fit lines among categories (*P* values for slope and intercept respectively: maternal age, .16, .15; fetal sex, .80, .81; and maternal diabetes, .45, .37).

DISCUSSION

We demonstrate that a 3D U-Net produced accurate automated segmentations and volumetric measurements of fetal intracranial and brain volumes on clinical 2D MRIs acquired in 3 planes. The U-Net functions across a wide range of GAs (19–38 weeks). Automated estimates of intracranial and brain volume closely approximated calculations of intracranial volume based on manual segmentations (Fig 3). This method has the potential to provide accurate volumetric data that can be incorporated into assessments of fetal neurologic development in a clinical population.

After validating the accuracy of volumetric measurements, the automated deep learning–based method was applied to 184 fetal brain MRIs with normal findings to develop a reference standard of normal intracranial and brain volumes across GAs (Fig 4). By comparing individuals' automated volumes with this reference standard, we were able to correctly identify most pathologic (microcephalic and macrocephalic) brain volumes in a few seconds, demonstrating the potential clinical utility of this approach.

Our method was trained on and applied to a population of fetal brain MRIs with clinically normal findings, as assessed by the absence of reported abnormalities on a pediatric neuroradiologist's report. The variability of resulting measurements may be higher for this clinical population than for a population of completely normal pregnancies recruited for a specific research study, in which mothers with comorbidities might be excluded. Prior studies have deployed atlas-based methods for fetal brain segmentation.[13-15] Although methods such as those of Jarvis et al,[16] provide grossly similar total intracranial volume measurements, small differences in normative values were noted, and further research will need to identify whether such differences may be due to measurement techniques, patient sampling differences, or other factors. The advantage of applying our method in a heterogeneous clinical population is that normative values are needed for this population of clinically normal fetuses, to which the clinical tool would ultimately be applied.

The results of the automated segmentation of fetal brain MRIs using our 3D U-Net are comparable with those of prior studies using 2D U-Nets to segment the fetal brain.[7,17] For example, Li et al[7] demonstrated similarly high Dice scores with a 2D U-Net after training on 212 fetal brains (23–38 weeks of age) for the purpose of building an atlas of 35 fetal brains in the Chinese population. However, this study did not provide normative values across the population. Our results are also in line with a recent report using a 3D U-Net for multicompartment fetal brain segmentation, demonstrating improved results compared with atlas-based techniques.[18] However, this new method requires a slice-to-volume reconstruction before the application of the U-Net, which can be difficult to implement robustly, requiring additional technical expertise. In contrast, our method functions on 2D images directly, combining multiple measurements in different planes of acquisition, simplifying the method's use. We make this method and all related code publicly available and easy to use so that larger normative data sets may be easily built across institutions.

The automated deep learning method developed here lends itself well to both clinical and research use. For clinical use, the speed of processing of the method allows near real-time quantitative volumetric data to be obtained. From a research perspective, the method can be used at the population level to examine associations between various genetic or environmental exposures and fetal brain development. For example, by using automated volume calculations from our normative population, we demonstrate that in our data, there are no significant effects of maternal diabetes,[19] fetal sex,[20] or maternal age[21] on the relationship between GA and total intracranial volume. The lack of an effect of fetal sex on brain
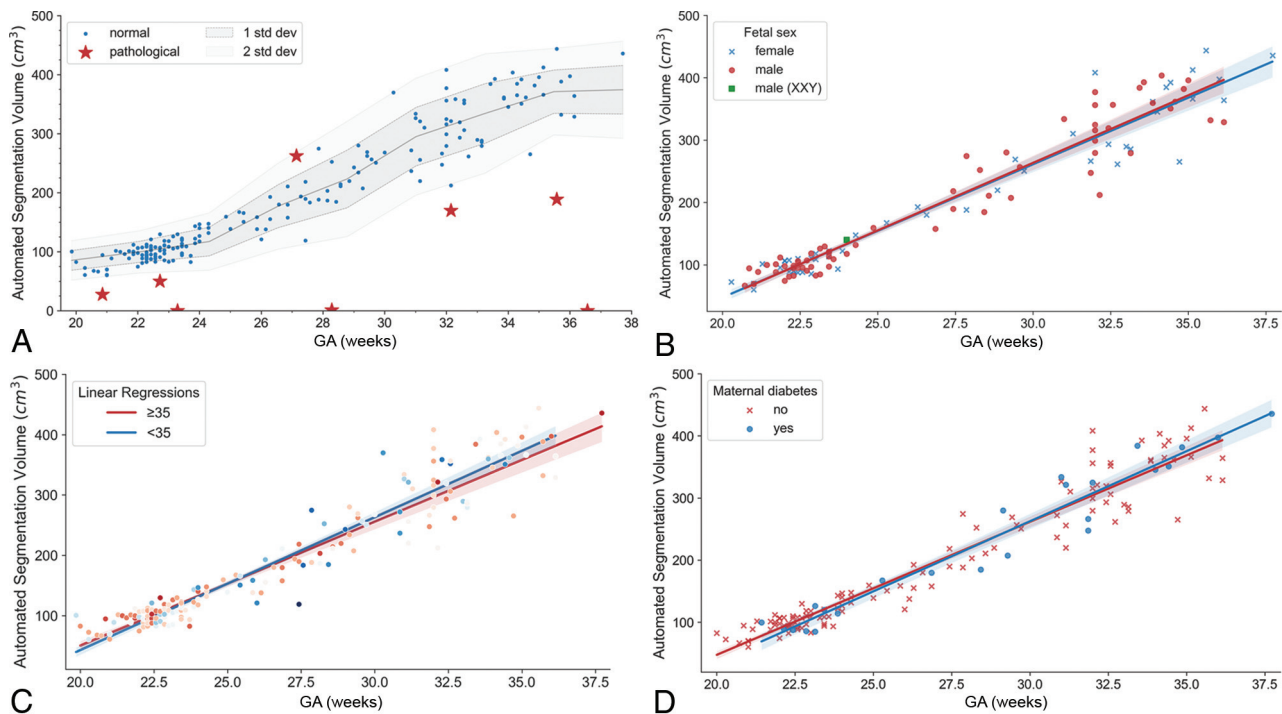
FIG 5. *A*, Intracranial volumes as a function of GA, as shown in Fig 4, with estimates of intracranial volume on pathologic brains added (*red markers*). Pathologic brain volumes fall outside the range of normal values from the normative data set. *B*, Intracranial volumes as a function of GA color-coded by fetal sex. *C*, Intracranial volumes as a function of GA color-coded by maternal age. *D*, Intracranial volumes as a function of GA color-coded by maternal diabetes status. *B–D*, The *solid lines* are linear regression models, with the *shaded areas* representing 95% confidence intervals.

volume may seem surprising, given recent[22] and prior[20] results demonstrating sexual dimorphisms in brain volume subregions, including total brain volume. However, associational studies, including our own, have relatively small sample sizes (less than thousands of samples), which can result in seemingly conflicting results,[23] lending urgency to the need for simple tools that can quantify aspects of fetal brain MR imaging across multiple institutions to create large normative data sets. Larger samples may reveal, for example, that these sexual differences emerge only later in gestation,[20] which smaller samples cannot adequately resolve.

There are several limitations to this study. The study was only applied to images from a single institution; inference of fetal intracranial volumes at other institutions would require further algorithmic evaluation and possibly additional training.[24] The algorithm was trained on only a small number of studies, and while this training combined with efficient use of data resulted in impressive performance, it is likely that more accurate and more extensive normative data could be created by increasing the training data set size. A larger data set would also give the opportunity to create a validation split that could be used to perform hyperparameter optimization and thus additionally improve the performance of the model. Furthermore, the algorithm trained here only produces 2 quantitative values, namely intracranial and brain volumes. Future work will use similar methods to produce additional quantitative measurements, such as volumes of brain substructures (eg, cerebellum, pons, corpus callosum) or measures of whole-brain morphology (eg, sulcation), toward a more complete quantitative fetal brain evaluation. Finally, the current process of applying the U-Net to

each plane of acquisition individually was developed to provide a trade-off between accurate segmentations and manual intervention. While 9 acquisitions (3 in each plane) are obtained for each fetal MR imaging at our institution, our method currently requires manual selection of the best acquisition for each plane of imaging to avoid inclusion of acquisitions with motion or other artifacts, mirroring routine clinical practice. Methods could be incorporated to automatically disregard acquisitions with excessive artifacts or motion and to include additional available acquisitions for more robust volume measurements.

CONCLUSIONS

Automated deep learning methods can achieve accurate segmentations of fetal brain MR imaging and provide accurate quantitative estimates of fetal intracranial and brain volumes across a wide range of GAs. This method, which is made available to the research community, allows the largely automated creation of normative references for clinical and research applications.

Disclosure forms provided by the authors are available with the full text and PDF of this article at www.ajnr.org.

REFERENCES

1. Reddy UM, Filly RA, Copel JA; Pregnancy and Perinatology Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Department of Health and Human Services, NIH. **Prenatal imaging: ultrasonography and magnetic resonance imaging.** *Obstet Gynecol* 2008;112:145–57 CrossRef Medline

2. Glenn OA, Barkovich AJ. **Magnetic resonance imaging of the fetal brain and spine: an increasingly important tool in prenatal diagnosis, Part 1.** *AJNR Am J Neuroradiol* 2006;27:1604–11 Medline

3. Girard NJ, Chaumoitre K. **The brain in the belly: what and how of fetal neuroimaging?** *J Magn Reson Imaging* 2012;36:788–804 CrossRef Medline

4. Glenn OA, Barkovich AJ. **Magnetic resonance imaging of the fetal brain and spine: an increasingly important tool in prenatal diagnosis: Part 2.** *AJNR Am J Neuroradiol* 2006;27:1807–14 Medline

5. Parazzini C, Righini A, Rustico M, et al. **Prenatal magnetic resonance imaging: brain normal linear biometric values below 24 gestational weeks.** *Neuroradiol* 2008;50:877–83 CrossRef Medline

6. Conte G, Milani S, Palumbo G, et al. **Prenatal brain MR imaging: reference linear biometric centiles between 20 and 24 gestational weeks.** *AJNR Am J Neuroradiol* 2018;39:963–67 CrossRef Medline

7. Li H, Yan G, Luo W, et al. **Mapping fetal brain development based on automated segmentation and 4D brain atlasing.** *Brain Struct Funct* 2021;226:1961–72 CrossRef Medline

8. Andescavage NN, Plessis AD, McCarter R, et al. **Complex trajectories of brain development in the healthy human fetus.** *Cereb Cortex (Cortex)* 2017;27:5274–83 CrossRef Medline

9. Gholipour A, Rollins CK, Velasco-Annis C, et al. **A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth.** *Sci Rep* 2017;7:476 CrossRef Medline

10. Duong MT, Rudie JD, Wang J, et al. **Convolutional neural network for automated FLAIR lesion segmentation on clinical brain MR imaging.** *AJNR Am J Neuroradiol* 2019;40:1282–90 CrossRef Medline

11. Rudie JD, Weiss DA, Colby JB, et al. **Three-dimensional U-Net convolutional neural network for detection and segmentation of intracranial metastases.** *Radiology Artif Intell* 2021;3:e200204 CrossRef Medline

12. Platt JC, Simard PY, Steinkraus D. **Best practices for convolutional neural networks.** 2003. https://www.researchgate.net/publication/2880624. Accessed August 20, 2021

13. Habas PA, Kim K, Rousseau F, et al. **Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses.** *Hum Brain Mapp* 2010;31:1348–58 CrossRef Medline

14. Serag A, Aljabar P, Ball G, et al. **Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression.** *Neuroimage* 2012;59:2255–65 CrossRef Medline

15. Gholipour A, Estroff JA, Barnewolt CE, et al. **Fetal brain volumetry through MRI volumetric reconstruction and segmentation.** *Int J Comput Assist Radiol Surg* 2011;6:329–39 CrossRef Medline

16. Jarvis DA, Finney CR, Griffiths PD. **Normative volume measurements of the fetal intra-cranial compartments using 3D volume in utero MR imaging.** *Eur Radiol* 2019;29:3488–95 CrossRef Medline

17. Khalili N, Lessmann N, Turk E, et al. **Automatic brain tissue segmentation in fetal MRI using convolutional neural networks.** *Magn Reson Imaging* 2019;64:77–89 CrossRef Medline

18. Zhao L, Asis-Cruz JD, Feng X, et al. **Automated 3D fetal brain segmentation using an optimized deep learning approach.** *AJNR Am J Neuroradiol* 2022;43:448–54 CrossRef Medline

19. Denison FC, Macnaught G, Semple SI, et al. **Brain development in fetuses of mothers with diabetes: a case-control MR imaging study.** *AJNR Am J Neuroradiol* 2017;38:1037–44 CrossRef Medline

20. Studholme C, Kroenke CD, Dighe M. **Motion corrected MRI differentiates male and female human brain growth trajectories from mid-gestation.** *Nat Commun* 2020;11:3038 CrossRef Medline

21. Knickmeyer RC, Xia K, Lu Z, et al. **Impact of demographic and obstetric factors on infant brain volumes: a population neuroscience study.** *Cereb Cortex* 2017;27:5616–25 CrossRef Medline

22. Machado-Rivas F, Gandhi J, Choi JJ, et al. **Normal growth, sexual dimorphism, and lateral asymmetries at fetal brain MRI.** *Radiology* 2022;303:162–70 CrossRef Medline

23. Marek S, Tervo-Clemmens B, Calabro FJ, et al. **Reproducible brain-wide association studies require thousands of individuals.** *Nature* 2022;603:654–60 CrossRef Medline

24. Rauschecker AM, Nedelec P, Sugrue LP, et al. **Interinstitutional portability of a deep learning brain MRI lesion segmentation algorithm.** *Radiol Artif Intell* 2022;4:e200152 CrossRef Medline