AJNR

Automatic Localization of the Pons and Vermis on Fetal Brain MR Imaging Using a U-Net Deep Learning Model

Farzan Vahedifard, Xuchu Liu, Jubril O. Adepoju, Shiqiao Zhao, H. Asher Ai, Kranthi K. Marathu, Mark Supanich, Sharon E. Byrd and Jie Deng

Automatic Localization of the Pons and Vermis on Fetal Brain MR Imaging Using a U-Net Deep Learning Model

Farzan Vahedifard, Xuchu Liu, Jubril O. Adepoju, Shiqiao Zhao, H. Asher Ai, Kranthi K. Marathu, Mark Supanich, Sharon E. Byrd, and Jie Deng

ABSTRACT

BACKGROUND AND PURPOSE: An MRI of the fetus can enhance the identification of perinatal developmental disorders, which improves the accuracy of ultrasound. Manual MRI measurements require training, time, and intra-variability concerns. Pediatric neuroradiologists are also in short supply. Our purpose was developing a deep learning model and pipeline for automatically identifying anatomic landmarks on the pons and vermis in fetal brain MR imaging and suggesting suitable images for measuring the pons and vermis.

MATERIALS AND METHODS: We retrospectively used 55 pregnant patients who underwent fetal brain MR imaging with a HASTE protocol. Pediatric neuroradiologists selected them for landmark annotation on sagittal single-shot T2-weighted images, and the clinically reliable method was used as the criterion standard for the measurement of the pons and vermis. A U-Net-based deep learning model was developed to automatically identify fetal brain anatomic landmarks, including the 2 anterior-posterior landmarks of the pons and 2 anterior-posterior and 2 superior-inferior landmarks of the vermis. Four-fold cross-validation was performed to test the accuracy of the model using randomly divided and sorted gestational age–divided data sets. A confidence score of model prediction was generated for each testing case.

RESULTS: Overall, 85% of the testing results showed a ≥90% confidence, with a mean error of <2.22 mm, providing overall better estimation results with fewer errors and higher confidence scores. The anterior and posterior pons and anterior vermis showed better estimation (which means fewer errors in landmark localization) and accuracy and a higher confidence level than other landmarks. We also developed a graphic user interface for clinical use.

CONCLUSIONS: This deep learning–facilitated pipeline practically shortens the time spent on selecting good-quality fetal brain images and performing anatomic measurements for radiologists.

ABBREVIATIONS: AI = artificial intelligence; AP = anterior-posterior; DL = deep learning; GA = gestational age; SI = superior-inferior

CNS abnormalities are relatively common in fetuses, ranging from 0.1% to 0.2% in live births and 3% to 6% in stillbirths.[1] A diagnosis of fetal brain abnormalities at an early stage is essential. Fetal sonography is considered the criterion standard of anatomic measurements. MR imaging is often performed when sonography is inconclusive to provide additional information for assessing fetal anatomy during all phases of gestation.[2] MR imaging provides superior soft-tissue contrast and spatial resolution for differentiating highly variable fetal brain tissue.[3] Fetal MR imaging combined with fetal sonography increases confidence in the early detection of perinatal disorders of development.

Manual measurements have several disadvantages, including clinicians' training requirements, time commitment, and inter- and intraobserver variability.[4] Radiologists must choose the highest quality image series without motion artifacts or missing anatomy and then identify and measure various anatomic structures.[5] Small measurement errors may result in misdiagnosis and misguided pregnancy management.[5] Accurate measurements of fetal brain anatomy are critical to differentiate hypoplastic, absent, or malformed brains from normal brain structures.[6] Measurement errors can have significant consequences in clinical practice because they can lead to misdiagnosis and misguided pregnancy management.

Fetal MR imaging interpretation also requires specialized training. However, there is a shortage of pediatric neuroradiologists, resulting in limited availability.

During the past decade, artificial intelligence (AI) algorithms, specifically deep learning (DL), have significantly advanced image-recognition tasks.[7] The machine learning approaches have the potential to aid in the early detection of these issues, thereby enhancing the diagnostic and follow-up processes.

By means of AI, processing of fetal brain MR imaging has investigated models that automatically predict specific landmarks and segmentation. Various AI models (primarily convolutional neural networks and U-Net) were used.[8,9] Some models achieved an accuracy of ≥95%. AI could aid in the pre- and postprocessing[10] and reconstruction,[11] predicting gestational age (GA) (with an accuracy of 1 week),[12] fetal brain extraction,[13] and fetal brain segmentation.[11,14] AI could also help in GA prediction, fetal motion detection, motion tracking, pose estimation,[15] and super-resolution reconstruction.[11] Recently, several publications developed AI models for automatic fetal brain anatomic measurements in a biparietal diameter,[16] which were derived from identified landmarks after several preprocessing steps such as computation of an ROI, reference section selection, segmentation, midsagittal line and fetal brain orientation, and, finally, measurements.

In this work, we focused on identifying 2 anterior-posterior (AP) landmarks of the pons and 2 AP and 2 superior-inferior (SI) landmarks of the vermis. All the landmarks on each structure were predicted simultaneously using U-Net multisegmentation features. We exploited U-Net[17] to determine imaging features surrounding a landmark point and calculated the probability of any image pixel being the defined landmark point.

In addition, the image pixel with the highest probability within the output Gaussian distribution mask was predicted as the landmark point by the U-Net model. Finally, we developed a tool that could be extended to clinical use on the basis of the prediction model to help radiologists select the best image series for interpretation and perform fetal brain anatomic measurements more efficiently.

MATERIALS AND METHODS
Database
This retrospective study, approved by the institutional review board with a waiver of consent, included 55 fetal MR imaging studies at different GAs. The studies were selected from a database of pregnant women who underwent routine clinical fetal screening at Rush University Medical Center, Chicago, Illinois, between 2007 and 2020. All the selected studies confirmed normal fetal brain development based on radiology reports. Expert pediatric neuroradiologists performed image-quality screening and landmark annotation on the exported sagittal T2-weighted HASTE images. Of these 55 patients, some patients had >1 image series, so the total number of image series was 100. We added image series for data augmentation, increasing the data set.

Six landmarks, including AP landmarks on the pons and AP/SI landmarks on the vermis (drawn by the radiologist), served as the ground truth. In addition, manual biometric measurements on the pons and vermis were performed according to the standard clinical recommendations.[18]

MR Imaging Protocol
Fetal MR images were obtained at our institution using Siemens 1.5T MR imaging scanners (Siemens, Erlangen, Germany), without sedation. Single-shot HASTE images were acquired in the axial, coronal, and sagittal planes with the following parameters: TR = 1400 ms, TE = 120 ms, FOV = 230×230 mm^2, and section thickness/gap = 3/0 mm, under free breathing. The fetal age range was 20–39 weeks, and the cases did not involve twins or significant maternal risk factors.

Image Preprocessing
All images were resized to 512×512 and augmented through rotating, flipping, adding Gaussian noise, motion blurring, median blurring, contrast-limited adaptive histogram equalization, sharpening, embossing, random brightness contrast adjustment, and random hue saturation adjustment using an open-source library (albumentations.ai; https://albumentations.ai/) with default parameter settings.

The study used original MR images without super-resolution reconstruction or other quality-selection preprocessing. A preselection process ensured suitable images for the study, focusing on landmark visibility and differentiation from neighboring structures. Exclusion criteria included motion artifacts, which hindered landmark identification and blurred anatomic borders. Consistency was maintained using midsagittal planes for pons and vermis measurements, avoiding oblique planes. Clinicians selected the data set for pons and vermis annotations, which underwent independent verification by radiologists and AI engineers for accuracy and consistency.

An innovative aspect of our research lies in the use of U-Net for landmark predictions. Rather than using the conventional binary segmentation output of 0/1, we modified the final output layer to generate a distribution map indicating the probability of landmark locations. This novel approach allowed us to extract valuable information from the U-Net model and precisely predict the positions of the landmarks, further enhancing the significance of our article.

Model Performance Validation
The U-Net model was used to fit the Gaussian distribution function. After the radiologist provided specific landmarks, the AI model calculated the distribution probability of these landmarks on the image by collecting the image features.

Model performance evaluation focuses on the probability relationship between image features and landmarks. Given an unsuitable image for labeling the vermis, AI will try to determine where the most likely landmark point will be. However, confidence in this point may be low due to differences in the features learned by image and AI.

With this probability distribution, we feed all the images into the pipeline and filter out the most suitable images for the physician's annotation.

U-Net Model as a Transforming Function
The encoding path of the U-Net model incorporates convolutional and max-pooling layers for feature extraction and dimension reduction, while the decoding path uses up-sampling and
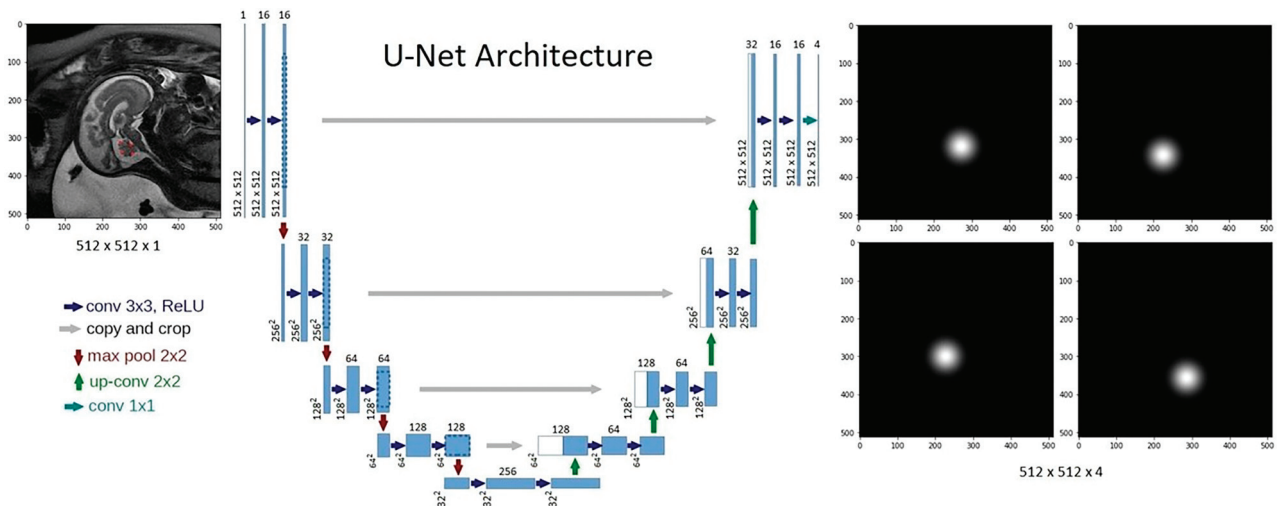
FIG 1. The architecture of the U-Net model for landmark detection. The U-Net converts a sagittal fetal brain image into gray-scale masks in which the vertex represents the locations of different landmark points. Up-conv indicates up-sampling operation; conv, convolutional layer; ReLU, Rectified Linear Activation.

concatenation for restoring spatial resolution and creating segmentation maps.

A U-Net model was built to transform an input MR image into a gray-scale image mask with its vertex representing the location of the predicted landmark point (Fig 1).

The transforming function $f(x, y)$ can be expressed as the impact of an arbitrary point (x, y) within the mask on the predicted landmark point (x_0, y_0) (Equation 1).

$$\text{Equation 1} \quad f(x, y) = f\left(\sqrt{(x - x_0)^2 + (y - y_0)^2}\right) = f(r),$$

where r represents the radial distance. The gray-scale mask is a rotationally symmetric function (radius of a circle: R) with a Gaussian distribution centered on the annotated landmark point (Fig 2), where R is proportional to the fronto-occipital radius of the fetal brain (Fig 3). We chose the Gaussian distribution function $f(r; \mu, \sigma)$ as $f(r)$ and simplified it by setting the mean and SD $\sigma = \frac{R/2}{\sqrt{2\pi}}$ as in Equation 2:

$$\text{Equation 2} \quad f(r; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r-\mu)^2}{2\sigma^2}} = \frac{2}{R} e^{-\frac{4\pi r^2}{R^2}}.$$

In practical applications, we removed the coefficient $\frac{2}{R}$ and replaced it in Equation 1.

$$\text{Equation 3} \quad f(r) = e^{-\frac{4\pi r^2}{R^2}}, \quad r \in [0, R]; \quad f(r) = 0, \quad r > R,$$

$$\frac{r}{R} \in [0, 1].$$

We customized the U-Net for landmark-prediction reliability by implementing a Gaussian output. This step permits AI to consider landmark-location uncertainty. The probability of a predicted point being the desired landmark was represented using this Gaussian function. The confidence score for each landmark was based on this probability. The model was trained to detect multiple landmarks simultaneously by generating separate masks representing the Gaussian probability of each landmark.

We balanced model complexity and computational efficiency by optimizing the hyperparameters of the model. We used a 3 × 3 kernel size to capture local contextual information efficiently. The channel depth gradually doubled after each max-pooling operation to learn more complex representations at different levels. The number of layers was chosen considering the task complexity and available computational resources, finding a suitable balance for the model.

Our study demonstrated that splitting the DL model outputs into separate models for pons and vermis landmarks resulted in improved accuracy. This approach allowed fine-tuned adjustments and enhanced detection of each landmark. By focusing on specific imaging features, the individual models improved the identification of anatomic structures. Compared with predicting all landmarks at once with a single model, this approach achieved superior performance and increased accuracy in detecting landmarks in fetal brain MR imaging.

Model Training

In the training process, a normalized weighted binary mean squared error loss function was used to compensate for the data imbalance. Other model parameters included batch size = 15 for both training and validation and epoch = 100 with an early stop after 10 epochs of loss increasing.

We used the Adam optimizer to update model weights efficiently, the EarlyStopping strategy to prevent overfitting, and the ModelCheckpoint callback to save weights of the best-performing model, ensuring reproducibility. The entire training process on the graphics processing unit Nvidia GeForce RTX 2080 Ti 11 Gb (NVIDIA GeForce RTX 2080 Ti 11Gb) took <2 hours.

K-Fold cross-validation

Two 4-fold cross-validation methods were implemented for model training and testing (Fig 4). In the first method, the data set was divided into 4 groups by a sorted range of GAs (ie, 20–26 weeks, 27–29 weeks, 30–33 weeks, and 34–39 weeks) (Fig 4A). In the second method, the data set was randomly divided with
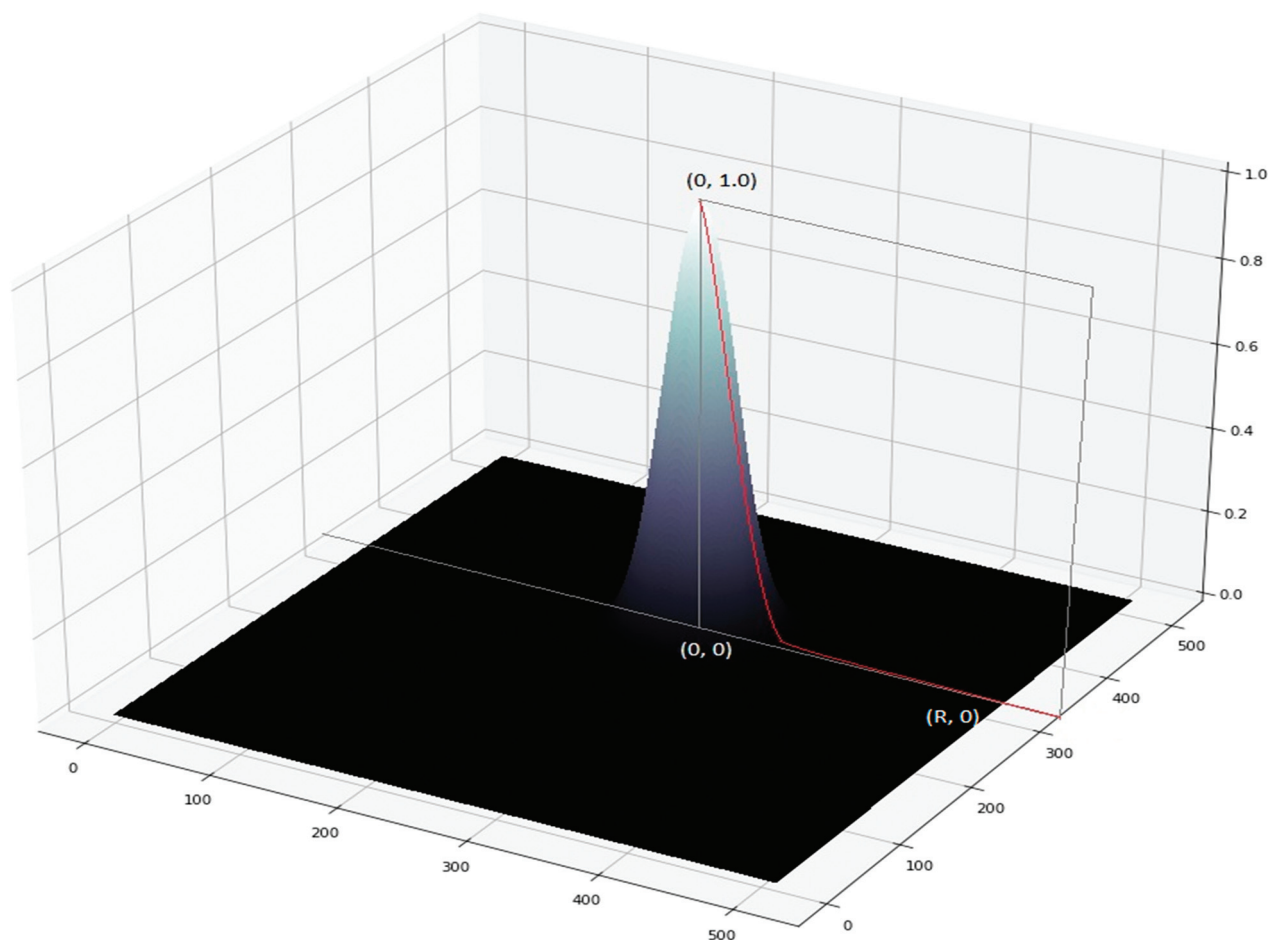
FIG 2. The cut surface of a 3D rotationally symmetric Gaussian distribution function with the radius (R), and the top represents the landmark point position.

mixed GAs without overlapping among all groups (Fig 4B). In each cross-validation fold, 3 groups of patient images were used as the training data set, and the other group, as the testing data set.

RESULTS

Model Performance

The randomly divided mixed GA method outperformed the sorted GA-divided method by providing smaller prediction errors ($P<.001$) with higher confidence scores ($P<.001$).

In the first cross-validation method, in which the data set was divided by sorted GA weeks, the prediction error distribution with associated confidence scores for 6 predicted landmarks in 100 image series (ie, the number of total landmarks: $6 \times 100 = 600$) is shown in Fig 5 (*A*, scatterplot; *B*, contour line plot). Among all 600 predicated landmark locations, 73% (440/600) showed a confidence score of >90%, with a mean prediction error of <2.25 mm. Among the 6 landmarks, the anterior vermis and anterior and posterior pons were best predicted with fewer errors and higher confidence scores (Online Supplemental Data). The anterior and posterior pons and anterior vermis had significantly fewer errors compared with the posterior vermis ($P<.01$, $P<.01$, and $P<.001$, respectively). Additionally, the anterior pons had a significantly higher confidence score than the superior/

inferior/posterior vermis ($P<.05$, $P<.05$, and $P<.01$, respectively), while the posterior pons had a significantly higher confidence score than the superior/posterior vermis ($P<.05$ and $P<.01$, respectively).

In the second cross-validation method, in which the data set was randomly divided with mixed GA weeks, the prediction error distribution with an associated confidence score for the 600 predicted landmarks is shown in Fig 6 (*A*, scatterplot; *B*, contour line plot). Among all 600 predicated landmarks, 85% (511/600) showed a confidence score of >90%, with a mean prediction error of <2.22 mm. Among the 6 landmarks, the posterior pons was the best-predicted landmark, with the smallest error and highest confidence score (Online Supplemental Data). The posterior pons had significantly lower error compared with all other landmarks ($P<.05$ for the anterior pons and anterior vermis, and $P<.001$ for the superior/inferior/posterior vermis). Additionally, the posterior pons had significantly higher confidence scores than the superior, inferior, and posterior vermis ($P<.05$, $P<.05$, and $P<.01$, respectively).

Automatic Landmark Detection

We evaluated the differences between manual landmark localization performed by a radiologist and an expert pediatric neuroradiologist, as shown in the Table. The variations between their
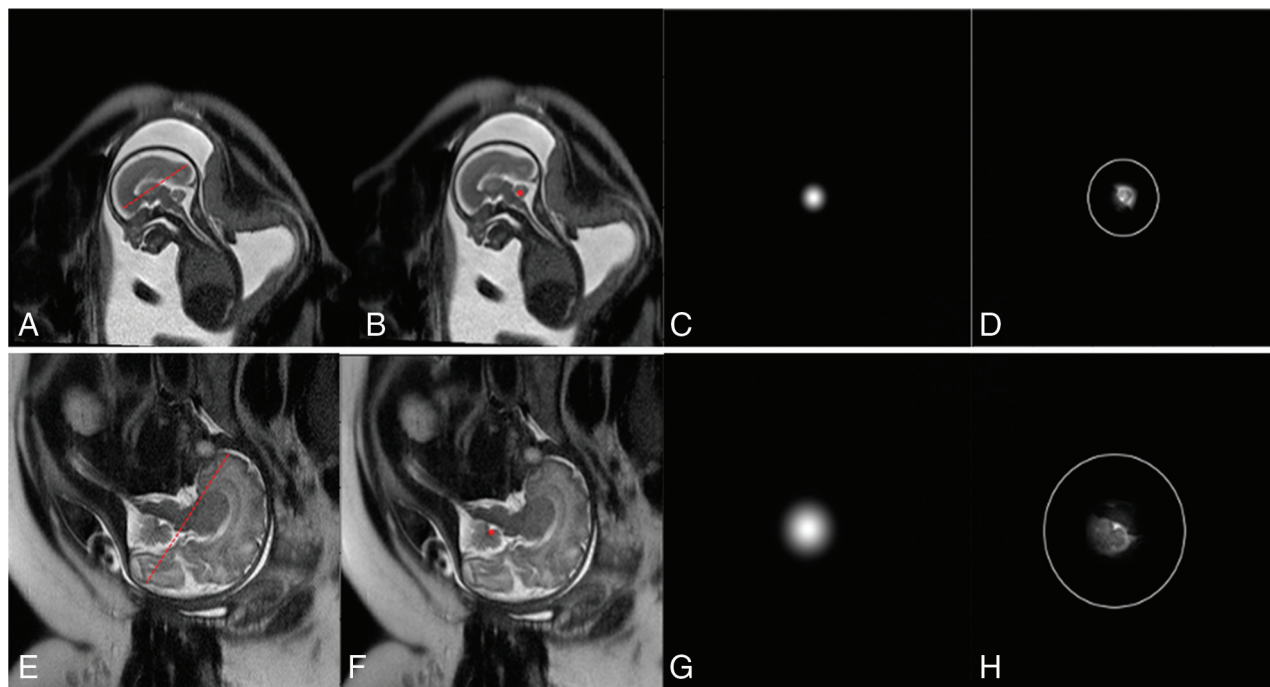
FIG 3. Fronto-occipital radius of the fetal brain determines the radius of Gaussian distribution function in a patient at GA week 20 (*A–D*) and another at GA week 33 (*E–H*). The left 2 columns are MR images with a fronto-occipital radius (*A* and *E*) and annotated landmark points on the vermis (*B* and *F*). The third column (*C* and *G*) shows the image mask with the Gaussian distribution used in model training. The fourth column shows the image area surrounding the landmark (*D* and *H*) determined by the Gaussian distribution function, with an added *white circle* indicating the radius range.

manual measurements ranged from a mean of 0.42 (SD, 0.59) mm for vermis1 to 1.87 (SD, 1.81) mm for vermis2. These disparities emphasize the presence of interrater variability and the possibility of measurement inconsistencies with manual assessments.

By automating landmark identification and using DL capabilities, our AI model consistently provides more reliable and consistent measurements.

Figure 7 shows examples of model-predicted landmark locations with biometric measurements between each pair demonstrated and compared with manual detections performed by the radiologist.

Despite variations in image quality and white noise levels, the AI system maintained accuracy, highlighting its robustness and adaptability. Figure 7 shows the precision and confidence of our AI model in 1 case. The left picture has a pixel spacing of 0.41, resulting in a relatively low resolution. The middle and right images have a pixel spacing of 0.35, which means better resolution. However, the right image has significantly more white noise points. The Online Supplemental Data show confidence and measurements in these 3 different series. The predictions of AI had an error rate below 0.5 mm and provided confidence scores for each prediction, enhancing trust in its outputs. (The AI provided confidence scores for each prediction and selected the optimal predictions with an error rate below 0.5 mm, thereby enhancing the trust of its outputs). Therefore, this AI model offers a reliable, accurate, and consistent tool for measurements.

Distance Measurement. The Online Supplemental Data illustrate the "distance measurements" of the pons and vermis, comparing

manual and AI measurements and the corresponding errors. The error-to-total measurement ratio is reasonable. For instance, when the confidence threshold is set above 90%, the pons distance from the average is 10.81 mm, with an error of 1.12 mm.

Statistical Analysis

Data were transformed to achieve a normal distribution before we conducted statistical testing. The prediction errors of the 6 landmarks and confidence scores were compared using paired *t* tests across the two 4-fold cross-validation methods. The Tukey Honest Significant Difference test was used to compare and rank the prediction errors and confidence scores for each landmark among all 6 landmarks. All statistical analyses were performed using R Studio (http://rstudio.org/download/desktop). A significance level of $P < .05$ was used.

Clinical Pipeline with Graphic User Interface

An interactive tool was developed on the basis of our developed model. This graphic user interface was created to help radiologists identify the landmarks of the pons and vermis and obtain biometric measurements on fetal MR imaging more efficiently (Fig 8). This graphic user interface will also provide confidence for each suggested image. To accept, reject, or modify the landmark prediction, the clinicians can judge according to the AI-provided confidence.

DISCUSSION

We proposed a novel U-Net DL model that automatically detects anatomical landmarks on the "pons" and "vermis" in fetal brain
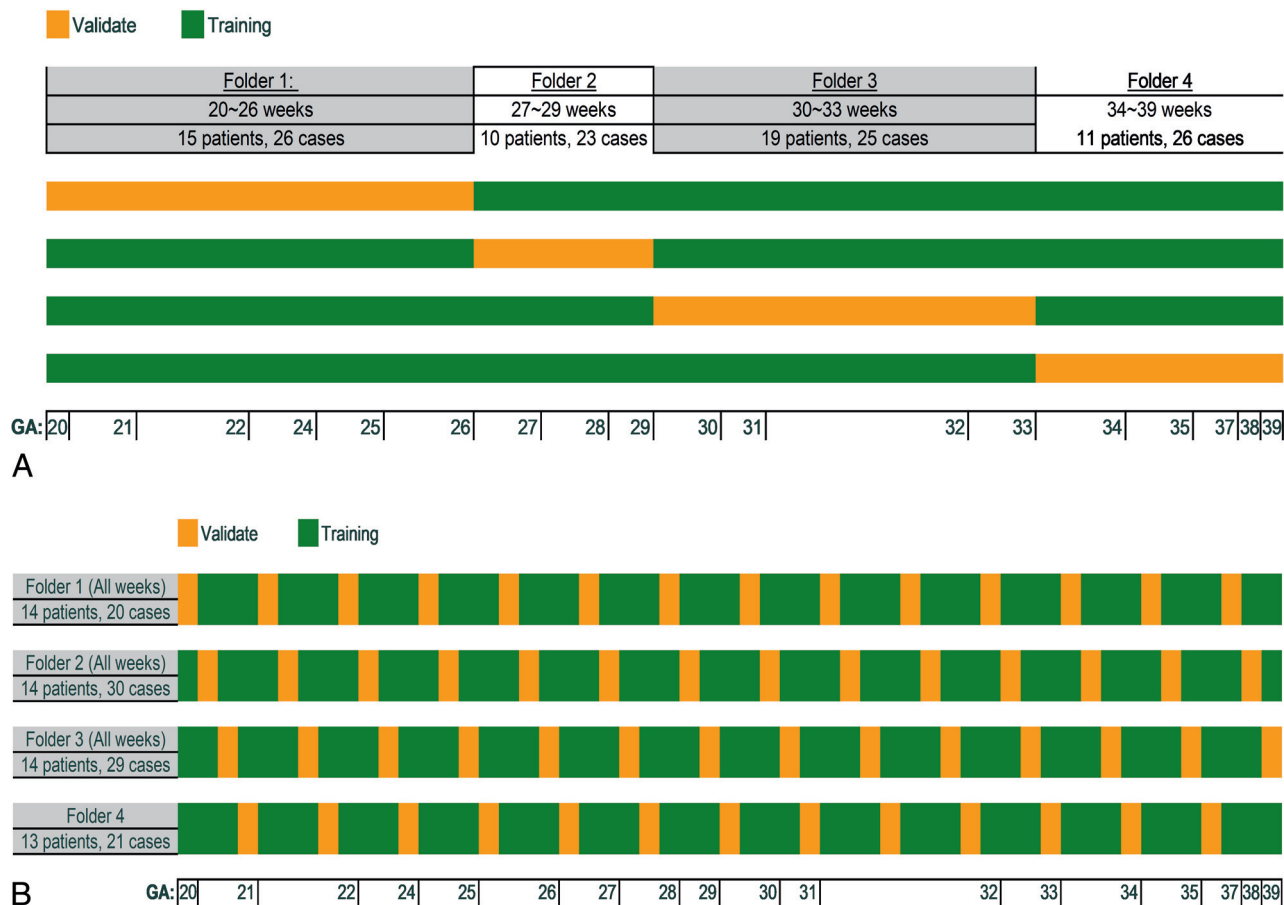
FIG 4. Two 4-fold cross-validation methods for DL-model training and testing. Method 1 divided the data sets by sorting the ranges of GA weeks (*A*). Method 2 divided the data sets randomly with mixed GA weeks (*B*).

MR imaging. Using these landmark locations, we predicted important fetal biometric parameters, including vermis diameter and height, and pons diameter. The critical component of our model was using U-Net as a transformational function to generate a gray-scale image mask with a Gaussian distribution by extracting the imaging features adjacent to the center of the mask. Although we only tested this model for landmarks on the pons and vermis with promising results, it may also be applied to detect landmarks on other brain structures with asymmetric image features.

This study initially used a 2-stage anisotropic 3D U-Net to detect fetal brain ROIs. A reference image section on which the landmarks were identified using a Fetal Measurement by Landmarks was used. A Gaussian Mixture Model estimated the landmark measurement reliability. Compared with fetal MR imaging radiologists, the model yielded a 95% confidence interval agreement of 3.70 mm for cerebral biparietal diameter, 2.20 mm for biparietal bone diameter, and 2.40 mm for transcerebral diameter. Our study is the first to automatically detect landmarks on the pons and vermis on fetal MR imaging using a U-Net as a function with very few training parameters, reducing the computational complexity and shortening the training time.

The accuracy of our model depends on learning imaging features around landmarks. Given the varying fetal brain size and appearance at different GAs, we set the mask size proportional to

the fronto-occipital diameter of each brain for adequate feature extraction. Among the 6 landmarks, 2 on the pons and 4 on the vermis, the anterior-posterior pons and anterior vermis had better accuracy, possibly due to distinctive adjacent image features like the fourth ventricle, aiding model learning.

We built a pipeline for automatic batch processing of multiple image series for landmark prediction. It selected the reference section on the basis of the highest confidence score and skipped poor-quality images. The reference section was presented to the radiologists for review, and they were alerted to manually adjust any landmark prediction with a confidence score of <0.8.

In model training and validation, we implemented 3 schemes for 4-fold cross-validation (dividing folds by GA weeks versus dividing folds by mixed weeks). The model trained and validated with mixed GA weeks provides overall better accuracy compared with the sorted GA week–divided approach, suggesting that a large number of imaging features extracted from fetal MR images with various white/gray matter contrast and anatomic details during fetal brain development are essential to include in model training.

In similar studies, Dovjak et al[19] conducted manual studies on cerebellar vermian lobulation and vermis/brainstem–specific markers using prenatal MR images. Their research improved hindbrain malformation classification and provided insights into vermian growth patterns. In contrast, our study introduces a DL model for automated identification and measurement of
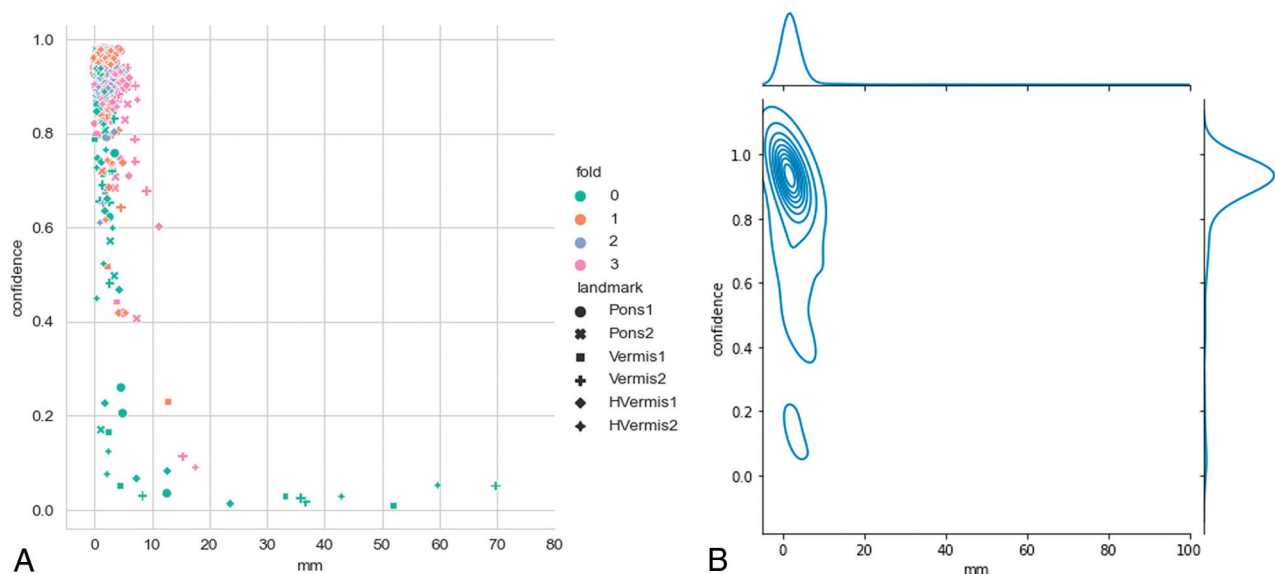
FIG 5. The scatterplot (*A*) and contour line plot (*B*) of the prediction error distribution with associated confidence scores in the sorted GA week validation method. The x-axis represents the distance (millimeters) between the predicted landmark and the ground truth, and the y-axis represents the confidence score of the prediction.
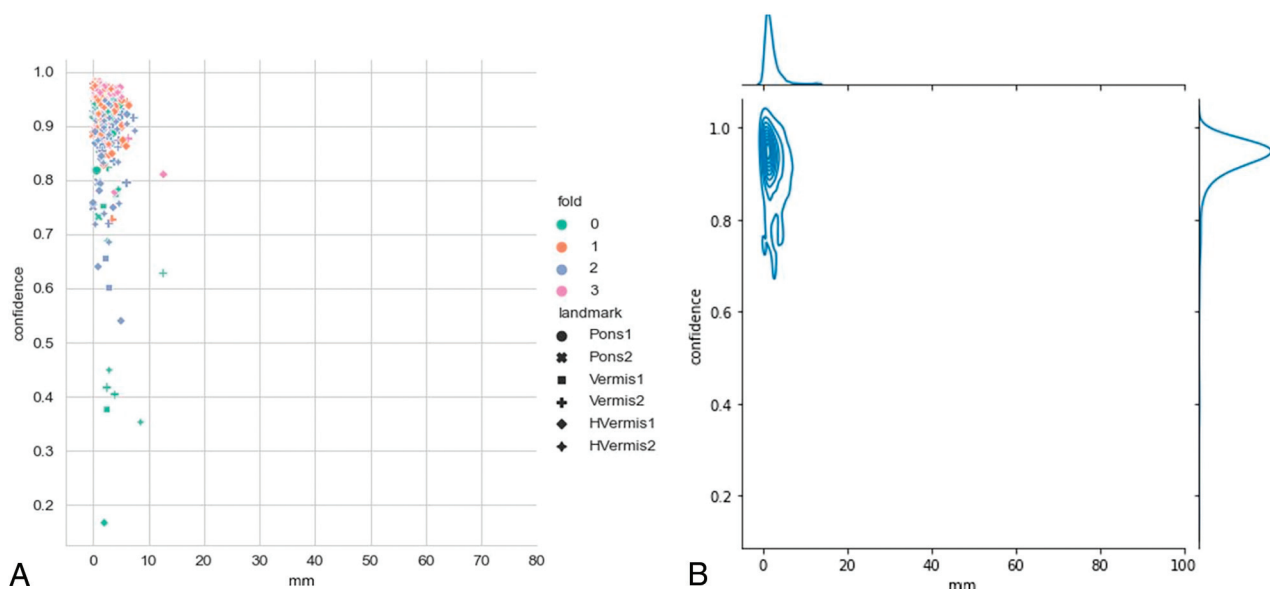
FIG 6. The scatterplot (*A*) and contour line plot (*B*) of the distribution of prediction error with associated confidence scores in the randomly mixed GA week validation method. The x-axis represents the distance (millimeters) between the predicted landmark and the ground truth, and the y-axis represents the confidence score of the prediction.

pons and vermis landmarks in fetal MR imaging. Our approach enhances efficiency, reduces errors, and offers confidence scores for predictions.

Advantages of Our Model

1) The benefit of automated landmark localization: The U-Net algorithm offers the advantage of automated landmark localization and provides associated confidence scores. This feature allows us to process all sagittal sequences of a patient and automatically identify the top 5 fetal brain MR image series with the highest confidence for physicians to choose from or manually adjust.

2) Time efficiency of the U-Net algorithm: Although the initial implementation and training of the U-Net algorithm require time and resources, its application on new fetal brain MR images is quick and efficient. The model can automatically identify landmarks and provide measurements without manual intervention. On our hardware (GeForce RTX 2080 Ti 12G), the average processing time for our U-Net model to predict 6 markers for an image is as low as 0.23 seconds, which is negligible compared with the time required by a physician for manual screening and measurements.

Compared manual landmark localization conducted by a radiologist and an expert pediatric neuroradiologist

	Pons1	Pons2	Vermis1	Vermis2	HVermis1	HVermis2
Mean (mm)	1.41	0.79	0.42	1.87	1.28	1.51
SD (mm)	1.12	0.76	0.59	1.81	1.38	1.68

Note:—Pons1 indicates Anterior landmark of Pons; Pons2, Posterior landmark of Pons; Vermis1, Anterior landmark of Vermis; Vermis2, Posterior landmark of Vermis; Hvermis1, Superior landmark of Vermis (height of vermis); Hvermis2, Inferior landmark of Vermis (height of vermis).
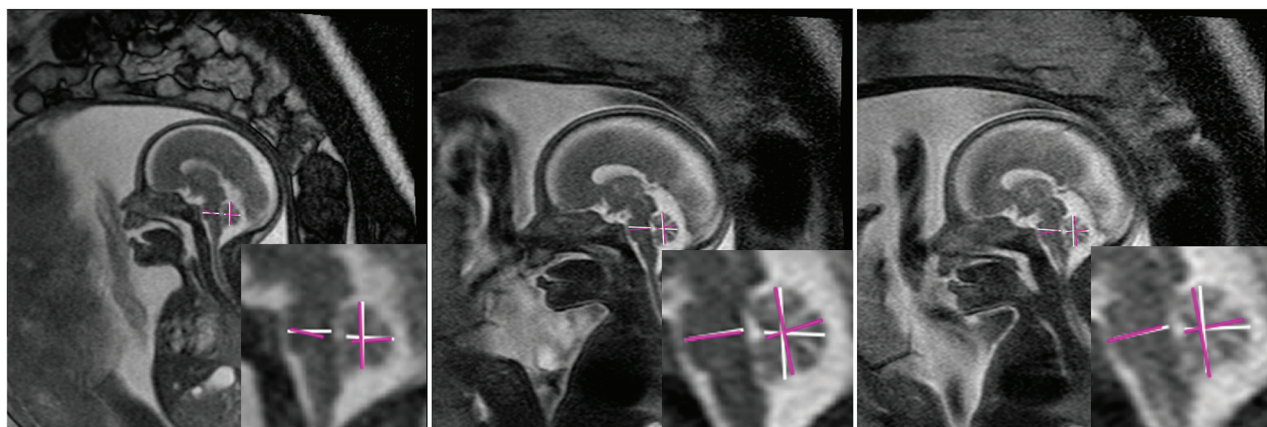
FIG 7. Representative image examples of model-predicted landmark locations with biometric measurements. (The *white line* is manual annotations by radiologists, and the *purple line* is DL-model predicted measurements). Three patients (*A*, 22 weeks; *B*, 22 weeks; *C*, 27 weeks) with accurate model-predicted landmarks compared with radiologists.

3) Consistency and reduced interobserver variability: The U-Net algorithm offers consistent landmark identification across different images and cases, reducing interobserver variability. This standardized approach leads to more reliable and reproducible measurements.

4) Accommodation of different resolutions and image quality: Our U-Net model robustly handles variations in resolution and evaluates image quality. It assigns confidence scores to each image slice, prioritizing high-quality images for accurate measurements and increased diagnostic confidence, resulting in errors of < 0.5 mm, regardless of resolution and image-quality differences.

5) Potential of the U-Net algorithm: The model demonstrates promising accuracy levels, potentially matching or exceeding human expert annotations. Further validation and optimization can enhance its reliability for posterior fossa biometry quantification.

6) Choosing linear measurements: Our study prioritized using linear measurements, specifically the diameter of the pons and vermis, due to their well-established clinical relevance and diagnostic utility. These measurements have been widely adopted in clinical practice and have demonstrated their effectiveness in detecting various brain abnormalities, including vermian hypoplasia, Dandy-Walker malformation, and pontocerebellar hypoplasia. Moreover, monitoring changes in the pons and vermis offers valuable insights into the neurologic development of the fetus, identifying potential issues and evaluating posterior fossa lesions.

The Run Algorithm on a Public Data Set

For an additional test, we ran the algorithm on the larger, publicly available Fetal Tissue Annotation and Segmentation Challenge (https://feta.grand-challenge.org/), resulting in good accuracy. Please see Online Supplemental Data, more codes and examples are provided.

Limitations

This study had some limitations: First, the sample size in each GA week range was small, possibly leading to insufficient model training. We used several strategies to enhance the accuracy and generalizability of our AI model to address the limitation of a limited number of cases. The MR imaging selection was conducted by experienced medical professionals, ensuring a high-quality training, validation, and testing data set. We used "transfer learning," enabling the model to identify distinctive features across a wider image range, thereby increasing its applicability in diverse clinical scenarios. Despite the size of our data set, the promising performance of the model in this pilot study suggests adaptability across different institutional settings. It provides landmark coordinates and confidence values, giving clinicians flexibility in MR image selection.

Second, this study did not assess fetal brain biometry in pathologic cases because of a paucity of cases with abnormal pons and vermis structures across different GA weeks. The study was limited to a single scanner platform and could potentially be restricted to a single cohort due to the limitation in the cohort. As a result, the generalizability of this study and its utility on large-scale data platforms may be limited. We recognize the need for further validation with larger, diverse data sets to ascertain the robustness and generalizability of our model in varied clinical environments.

Ensemble learning is a valuable option for future studies if additional data are available. This technique involves using all models obtained through 4-fold cross-validation and selecting
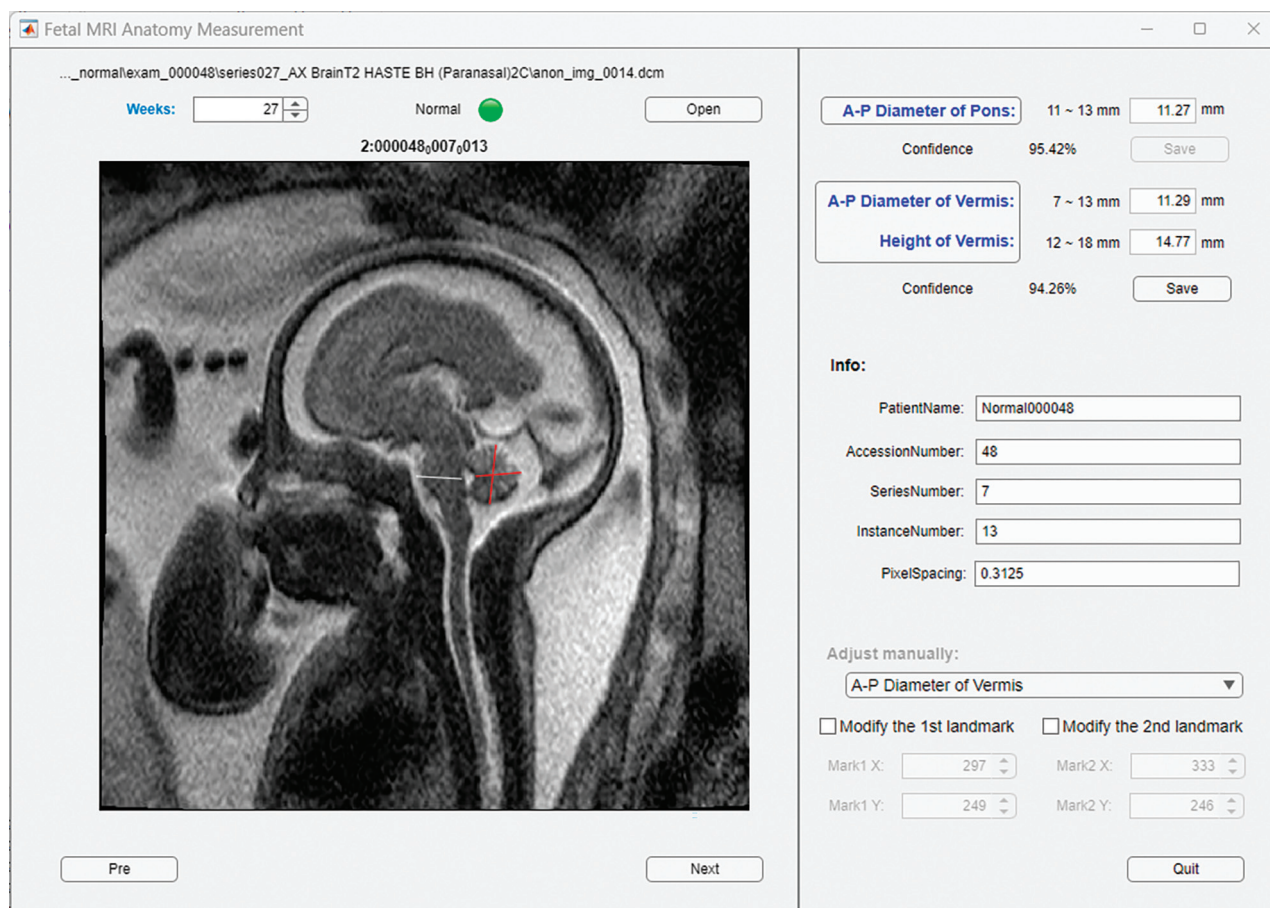
FIG 8. An interactive tool integrating DL-model-based prediction of landmarks. It helps the radiologist quickly locate, confirm, or adjust the landmarks on the automatically selected image section. After the landmark locations are confirmed, the distance between 2 related landmarks on a particular brain structure is calculated. The white line is the "Pons" AP diameter, and red crossed lines are "Vermis" AP diameter and height.

the landmark with the highest confidence, though it comes at the cost of increased processing time.

CONCLUSIONS

A U-Net model was developed to detect AP landmarks on the pons and AP/SI landmarks on the vermis. Our pipeline includes image series screening and selection, landmark prediction that improves radiologists' efficiency and time in identifying landmarks, performing anatomic measurements, and screening high-quality images. Using a U-Net-based DL model, we achieved a mean error of <2.22 mm and a ≥90% confidence score in 85% of the testing cases, resulting in improved estimation accuracy with reduced errors.

While manual measurements by radiologists often yield robust results, our AI model brings value by significantly reducing interrater variability and measurement errors. It accurately identifies high-confidence landmarks and optimizes image selection, even in instances in which blurred margins due to motion artifacts are present.

We also established a pipeline, graphic user interface, consisting of imaging selection and landmark prediction, followed by an interactive second check tool to help radiologists quickly locate, confirm, or adjust the landmarks on the autoselected image slices.

Using U-Net as a transformational function, our model accurately extracts imaging features around landmarks, particularly for the anterior and posterior pons and anterior vermis.

We validated this algorithm on a public data set, demonstrating good accuracy. The AI model addresses interrater variability, reduces measurement errors, and saves time, presenting advantages over manual measurements. Implementing AI-driven automation enables comprehensive and efficient fetal brain MR imaging assessment, potentially enhancing radiologists' efficiency and diagnostic accuracy and improving patient outcomes in fetal brain MR imaging analysis.

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

REFERENCES

1. Cater SW, Boyd BK, Ghate SV. **Abnormalities of the fetal central nervous system: prenatal US diagnosis with postnatal correlation.** *Radiographics* 2020;40:1458–72 CrossRef Medline
2. Hibbeln JF, Shors SM, Byrd SE. **MRI: is there a role in obstetrics?** *Clin Obstet Gynecol* 2012;55:352–66 CrossRef Medline
3. Shi Y, Xue Y, Chen C, et al. **Association of gestational age with MRI-based biometrics of brain development in fetuses.** *BMC Med Imaging* 2020;20:125 CrossRef Medline

4. Joskowicz L, Cohen D, Caplan N, et al. **Inter-observer variability of manual contour delineation of structures in CT.** *Eur Radiol* 2019;29:1391–99 CrossRef Medline

5. Warrander LK, Ingram E, Heazell AE, et al. **Evaluating the accuracy and precision of sonographic fetal weight estimation models in extremely early-onset fetal growth restriction.** *Acta Obstet Gynecol Scand* 2020;99:364–73 CrossRef

6. Rajagopalan V, Deoni S, Panigrahy A, et al. **Is fetal MRI ready for neuroimaging prime time? An examination of progress and remaining areas for development.** *Dev Cogn Neurosci* 2021;51:100999 CrossRef Medline

7. Hosny A, Parmar C, Quackenbush J, et al. **Artificial intelligence in radiology.** *Nat Rev Cancer* 2018;18:500–10 CrossRef Medline

8. Vahedifard F, Adepoju JO, Supanich M, et al. **Review of deep learning and artificial intelligence models in fetal brain magnetic resonance imaging.** *World J Clin Cases* 2023;11:3725–35 CrossRef Medline

9. Vahedifard F, Ai HA, Supanich MP, et al. **Automatic ventriculomegaly detection in fetal brain MRI: A step-by-step deep learning model for novel 2D-3D linear measurements.** *Diagnostics* 2023;13:2355 CrossRef Medline

10. Chen Z, Pawar K, Ekanayake M, et al. **Deep learning for image enhancement and correction in magnetic resonance imaging: state-of-the-art and challenges.** *J Digit Imaging* 2023;36:204–30 CrossRef Medline

11. Ebner M, Wang G, Li W, et al. **An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI.** *Neuroimage* 2020;206:116324 CrossRef Medline

12. Shen L, Zheng J, Lee EH, et al. **Attention-guided deep learning for gestational age prediction using fetal brain MRI.** *Sci Rep* 2022;12:1408 CrossRef Medline

13. Išgum I, Benders MJ, Avants B, et al. **Evaluation of automatic neonatal brain segmentation algorithms: the NeoBrainS12 challenge.** *Med Image Anal* 2015;20:135–51 CrossRef Medline

14. Makropoulos A, Counsell SJ, Rueckert D. **A review on automatic fetal and neonatal brain MRI segmentation.** *Neuroimage* 2018;170:231–48 CrossRef Medline

15. Gagoski B, Xu J, Wighton P, et al. **Automated detection and reacquisition of motion-degraded images in fetal HASTE imaging at 3 T.** *Magn Reson Med* 2022;87:1914–22 CrossRef Medline

16. Avisdris N, Yehuda B, Ben-Zvi O, et al. **Automatic linear measurements of the fetal brain on MRI with deep neural networks.** *Int J Comput Assist Radiol Surg* 2021;16:1481–92 CrossRef Medline

17. Ronneberger O, Fischer P, Brox T, eds. **U-Net: convolutional networks for biomedical image segmentation.** In: Navab N, Hornegger J, Wells WM, et al, eds. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015.* Springer Link; 2015

18. Garel C, Delezoide AL. *MRI of the Fetal Brain: Normal Development and Cerebral Pathologies.* Springer-Verlag; 2004

19. Dovjak GO, Diogo MC, Brugger PC, et al. **Quantitative fetal magnetic resonance imaging assessment of cystic posterior fossa malformations.** *Ultrasound Obstet Gynecol* 2020;56:78–85 CrossRef Medline