Deep learning model for predicting gestational age after the first trimester using fetal MRI

Kojita, Yasuyuki ; Matsuo, Hidetoshi ; Kanda, Tomonori ; Nishio, Mizuho ; Sofue, Keitaro ; Nogami, Munenobu ; Kono, Atsushi K. ; Hori,…

(Citation)
European Radiology,31(6):3775-3782

(Issue Date)
2021-06

(Resource Type)
journal article

(Version)
Accepted Manuscript

Deep learning model for predicting gestational age after the first trimester using fetal MRI

Yasuyuki Kojita, MD., Hidetoshi Matsuo, MD., Tomonori Kanda, MD. Ph.D., Mizuho Nishio, MD. Ph.D., Keitaro Sofue, MD. Ph.D., Munenobu Nogami, MD. Ph.D., Atsushi K Kono, MD. Ph.D., Masatoshi Hori, MD. Ph.D., Takamichi Murakami, MD. Ph.D.

Department of Radiology, Kobe University School of Medicine, Kobe, Japan

Corresponding author: Yasuyuki Kojita

Department of Radiology, Kobe University School of Medicine,

7-5-2 Kusunoki-cho, Chuo-ku, Kobe, Hyogo, 650-0017, Japan

Tel.: + 81-78-382-6104

Fax: + 81-78-382-6129

E-mail: y.kojita10312@gmail.com

Type of article: Original Research

Deep learning model for predicting gestational age after the first trimester using fetal MRI

Original Research

Abstract

Objectives: To evaluate a deep learning model for predicting gestational age from fetal brain MRI acquired after the first trimester in comparison to biparietal diameter (BPD).

Materials and methods: Our Institutional Review Board approved this retrospective study, and a total of 184 T2-weighted MRI acquisitions from 184 fetuses (mean gestational age: 29.4 weeks) who underwent MRI between January 2014 and June 2019 were included. The reference standard gestational age was based on the last menstruation and ultrasonography measurements in the first trimester. The deep learning model was trained with T2-weighted images from 126 training cases and 29 validation cases. The remaining 29 cases were used as test data, with fetal age estimated by both the model and BPD measurement. The relationship between the estimated gestational age and the reference standard was evaluated with Lin's concordance correlation coefficient (ρc) and a Bland-Altman plot. The ρc was assessed with McBride's definition.

Results: The ρc of the model prediction was substantial (ρc = 0.964), but the ρc of the BPD prediction was moderate (ρc = 0.920). Both the model and BPD predictions had greater differences from the reference standard at increasing gestational age. However, the upper limit of the model's prediction (2.45 weeks) was significantly shorter than that of BPD (5.62 weeks).

Conclusions: Deep learning can accurately predict gestational age from fetal brain MR acquired after the first trimester.

Keywords

Gestational age; Brain; Fetus; Pregnancy; Deep learning

Key points:

- The prediction of gestational age using ultrasound is accurate in the first trimester but becomes inaccurate as gestational age increases.
- Deep learning can accurately predict gestational age from fetal brain MRI acquired in the second and third trimester.
- Prediction of gestational age by deep learning may have benefits for prenatal care in pregnancies that are underserved during the first trimester.

List of all abbreviations

BPD: biparietal diameter, CNN: convolutional neural network, CI: confidence interval, ρc: Lin's concordance correlation coefficient, DICOM: Digital Imaging and Communications in Medicine

Introduction

Accurate prediction of gestational age is essential for determining the appropriate treatment and management of pregnancies [1]. The last menstruation and ultrasonography findings are usually used to predict gestational age [2,3], with determination of gestational age on ultrasonography being based on head circumference, abdominal circumference, biparietal diameter (BPD), and femoral length. The prediction of gestational age using ultrasound is accurate in the first trimester but becomes inaccurate as gestational age increases [1]. It has been reported that the error range in the first trimester is within 1 week, but that it increases to 1–2 weeks in the second trimester and 3 weeks to 1 month in the third trimester [1]. Accurate determination of gestational age is fundamental to obstetric care and can have a positive effect on pregnancy outcomes. In a previous study, women receiving routine ultrasonography during the first trimester had significantly less need for post-term inductions than women receiving ultrasonography during only the second trimester [4]. Prediction of gestational age by ultrasound after the first trimester may overestimate the gestational age and causes unnecessary post-term inductions. Therefore, there is a need for an accurate method of estimating gestational age after the first trimester.

Deep learning is a type of machine learning that uses a computational model composed of multiple processing layers to enable automated classification of items in a dataset [5,6]. Convolutional neural networks (CNNs) are a commonly used deep learning method for performing image analysis, and they are suitable for medical image recognition tasks [7–12]. CNNs have been shown to be able to accurately predict the age of adult subjects from brain MRI [13,14].

As the shape of the fetal brain changes with gestational age [15], we hypothesized that gestational age could be predicted from fetal brain MRI using a CNN-based deep learning model than using BPD. The purpose of this study was to evaluate the use of a deep learning model for determining gestational age from fetal brain MRI acquired after the first trimester in comparison with BPD.

Materials and Methods

Study Participants

Our institutional review board approved this study and waived the requirement for written informed consent because of its retrospective nature. Two hundred twenty-five consecutive fetal MRI acquisitions obtained for imaging diagnosis at our institution between January 2014 and June 2019 were included. The exclusion criteria were a lack of clinical data, fetuses in the first trimester, fetuses with congenital malformation before and after birth, and the examination's status as a second or subsequent MRI. If a fetus underwent multiple MRI examinations, only the images obtained during the first MRI examination were used. A flow chart summarizing the inclusion and exclusion criteria is presented in Fig. 1. The reference standard gestational age was estimated from the last menstruation date and the ultrasonography measurements in the first trimester performed by the pregnant mother's referring obstetrician. The reasons why the clinical fetal MRI scans were requested are given in Table 1. The fetal MRI dataset was divided into three separate subsets: a training dataset used to optimize the model parameters, a validation dataset used to tune the model hyperparameters, and a test dataset used to evaluate the performance of the model relative to the reference standard. Two-thirds of the dataset were used as training data, one-sixth as validation data, and one-sixth as test data, with similar age distributions present in each group.

Image Preprocessing

The images were acquired with multiple platforms, as follows: 170 images on Achieva (1.5T, Philips Healthcare), 6 images on SIGNA EXCITE (1.5T, GE Healthcare), 3 images on Ingenia (3T, Philips Healthcare), 2 images on Signa Voyager (1.5T, GE Healthcare), 1 image on Intera (1.5T, Philips Healthcare), 1 image on Avanto (1.5T, Siemens Healthineers), and 1 image on SymphonyTim (1.5T, Siemens Healthineers). Image evaluation was performed on axial single-shot fast spin echo T2-weighted images of the fetal brain at the level of the thalamus and septum pellucidum, with the imaging parameters including a slice thickness of 2–8 mm, a matrix size of 256 × 256 to 640 × 640, and a median field of view of 400 mm (range, 150–600 mm). If the image was oblique or the anatomical structure was unclear because of fetal movement, the image closest to the level of the thalamus and septum pellucidum was selected (with reference to the shape of the lateral ventricles). The percentage of images that were oblique or contained artifacts was also recorded. For the deep learning, the images were converted from Digital Imaging and Communications in Medicine (DICOM) format to Tagged Image File format images of 512 × 512 pixels. Then, LabelImg software (ver. 1.8.3; https://github.com/tzutalin/labelImg) was used to crop a square area around the fetal head at the level of the thalamus and septum pellucidum, and the cropped images were scaled to 150 × 150 pixels by the nearest neighbor resizing method. The median x-axis value per pixel was 0.60 mm (range, 0.18–0.81 mm), and the median y-axis value was 0.56 mm (range, 0.14–1.17 mm). For the test dataset, BPD was measured on MRI in the same way as it was measured on ultrasonography, with BPD defined as the diameter from the outer edge of the calvaria to the inner edge of the opposite side of the calvaria on the axial plane traversing the thalamus and septum pellucidum. The conversion table used to calculate gestational age from BPD measured on ultrasonography examinations was also applied to the BPD measurements made on MRI [16].

Model Implementation

In this study, we developed a deep learning regression model that used fetal MRI as the input data and predicted gestational age (Fig. 2). To improve generalizability, VGG16, a discrimination model pre-trained with images used in the Large Scale Visual Recognition Challenge 2014, was modified for regression and used for transfer learning [17,18]. All processing was performed on a PC without a discrete graphics processing unit (CPU: Ryzen 5 Pro 3500U 2.1 GHz, 16 GB RAM). Python (3.7.6) (http://www.python.org) was used as the programming language, and Keras 2.2.4 (https://github.com/fchollet/keras) and TensorFlow

1.15.4 (http://tensorflow.org/) were used as the deep learning framework. To normalize the data values, the input images' pixel values were treated as integer values of 0–256 inside the program, and they were divided by 256 to convert them to continuous values of 0–1. The output (weeks) was also divided by 40. The input layer of the pre-trained VGG16 was resized to 150 × 150 pixels to match the size of the fetal MR images, the sigmoid layer just before the output layer was changed to a linear activation layer, and the output dimension was changed to 1. Transfer learning was performed with freezing of the trainable parameters in 13 layers of the VGG16 network, including the convolutional layer. Mean absolute error was used as the regression loss function, and stochastic gradient descent was used as the optimizer (learning rate = 0.0001). The modified VGG16 network was trained using a batch size of 1 and up to 50 epochs. Training was completed when the loss on the validation set did not improve, and predictions were then performed on the test data after fixing the model. In every epoch, 126 images were used for the training. These images were augmented by rotations (−180°–180°), vertical or horizontal flipping, and 0.8–1.2× zooming to avoid overfitting during training [8]. Finally, the trained model was used to predict the gestational age of the fetuses in the test dataset, and the results were denormalized.

Statistical Analysis

The gestational age as predicted by the model or BPD was compared with the reference standard using Lin's concordance correlation coefficient (ρc) and assessed according to McBride's definition, as follows: less than 0.9, poor; 0.90–0.95, moderate; greater than 0.95–0.99, substantial; greater than 0.99, almost perfect [19]. We used Lin's concordance correlation coefficient to evaluate the concordance according to previous studies [19-22]. To evaluate any systematic bias, the gestational age predicted by the model or BPD was compared with the reference standard using a Bland-Altman plot. All analyses were conducted with MEDCALC (version 19; https://www.medcalc.org/).

Results

A total of 184 T2-weighted MRI acquisitions from 184 fetuses were included. The mean fetal gestational age was 29.4 weeks (range, 14.0–41.4 weeks), and 25% (n = 46) and 75% (n = 138) were in the second and third trimesters, respectively. The distribution of gestational age is shown in Fig. 3. There were no apparent anomalies in all 184 fetuses. The 184 images were randomly divided into a training dataset of 126 images (68.4%), a validation dataset of 29 images (15.8%), and a test dataset of 29 images (15.8%). The gestational age distribution was very similar across the datasets (Table 2). Fig. 4 shows the typical trimming of fetal MR images and the results of the model. Out of the 184 total images, 100 (54.3%) were oblique and 31 (16.8%) contained artifacts, with 23 images (12.5%) both being oblique and containing artifacts, resulting in an overall total of 108 images (58.7%) that were oblique and/or contained artifacts. The ρc values for the agreement in predicted age with the reference standard are shown in Fig. 5. The agreement between the model prediction and the reference standard was substantial (ρc = 0.964, 95% confidence interval [CI]: 0.928–0.982), whereas the agreement between the BPD prediction and the reference standard was moderate (ρc = 0.920, 95% CI: 0.845–0.959), although the 95% CIs overlapped. Bland-Altman plots illustrating the difference between the model or BPD prediction and the reference standard are shown in Fig. 6. The model prediction was 0.43 ± 1.47 (mean ± standard deviation, 95% CI: 0.13–0.99) weeks shorter than the reference standard, whereas the BPD prediction was 0.97 ± 2.37 (95% CI: 0.07–1.87) weeks longer. Compared with the model predictions, BPD predictions tended to show larger errors with increasing gestational age. The lower limit of agreement between the model prediction and the reference standard (−3.31, 95% CI: −4.27 to −2.34) was almost the same as the lower limit of agreement between the BPD prediction and the reference standard (−3.67, 95% CI: −5.23 to −2.11). The upper limit of agreement between the model prediction and the reference standard (2.45, 95% CI: 1.48–3.41) was smaller than the upper limit of agreement between the BPD prediction and the reference standard (5.61, 95% CI: 4.06–7.18).

Discussion

Gestational age prediction is essential for perinatal management, but the prediction of gestational age by ultrasonography is not accurate after the first trimester [1]. We predicted the gestational age from fetal brain MRI after the first trimester using a deep learning model and BPD. The agreement between the model prediction and the reference standard was substantial, but the agreement between the BPD prediction and the reference standard was moderate. These results suggest that the deep learning method could be useful for predicting gestational age from fetal brain MRI, and that the method may be useful for evaluation of fetal development.

Wu et al. predicted gestational age in the third trimester by measuring cortical folding on fetal MRI and found a mean error of 0.43 weeks [23]. However, their computation of three-dimensional measurements required detailed brain segmentation and was weak in relation to motion artifacts such as fetal movement. In our study, we used T2-weighted images routinely acquired on fetal MRI and did not need additional images. Although 58.7% of the images were oblique or contained artifacts, the prediction of gestational age was accurate. The prediction of gestational age from fetal brain MRI with deep learning has the advantages that there is no need for additional images and that it is robust against fetal movement. Although there can be difficulties ensuring adequate quality in fetal MRI because of fetal movement and time constraints, the relatively high accuracy of predictions under such conditions demonstrates the versatility of deep learning. In addition, gestational age could be predicted not only in the third trimester, but also in the second trimester. Recently, motion-corrected fetal brain reconstruction methods have been used [24]. Using these methods may improve the image quality and the accuracy of deep learning evaluations.

The difference between the reference standard and the gestational age predicted by BPD increased with increasing gestational age. As can be seen in Fig 6, the model prediction was 0.43 (95% CI: −0.13–0.99) weeks shorter than the reference standard, whereas the BPD prediction was 0.97 (95% CI: 0.07–1.87) weeks longer. The standard deviation of the differences between the model and the reference standard was

1.47 weeks, which was smaller than the 2.37 weeks of BPD. The lower limit of agreement between the model prediction and the reference standard was −3.30 (95% CI: −4.27 to −2.34) weeks, which was almost the same as that for BPD: −3.67 (95% CI: −5.23 to −2.11) weeks. In contrast, the upper limit of agreement between the model prediction and the reference standard was 2.45 (95% CI: 1.48–3.41) weeks, which was smaller than that for BPD: 5.62 (95% CI: 4.06–7.18) weeks. Therefore, the model prediction has less variability and may reduce the risk of inappropriate pregnancy management.

In this study, we excluded images acquired in the first trimester. In the first trimester, it is often difficult to identify the head on fetal MRI, and gestational age can be predicted with sufficient accuracy by ultrasonography. However, as gestational age increases, it becomes more difficult to predict the gestational age by ultrasound. Deep learning can predict the gestational age more accurately from fetal brain MR acquired after the first trimester. Although such cases are very rare, for those who are not screened for gestational age during the first trimester and are only tested during the second or third trimester, deep learning may provide better obstetric care for pregnancy. In addition, the prediction of gestational age using deep learning is a different and a novel indicator of fetal development. Ultrasound is the first choice for fetal evaluation, and MRI should be performed in addition to ultrasound [25,26]. The prediction of gestational age using MRI is currently not considered as standard practice [26]. However, the findings of this study indicate that deep learning using MRI in the evaluation of gestational age has great future potential. Regarding the safety of MRI, in a large-scale epidemiological study in Canada, abnormalities did not occur in fetuses in the first, second, or third trimesters [27].

Our study has several limitations. First, the deep learning model used in this study was trained and tested using only a small number of images. The same combination of scanner manufacturer and magnetic field strength was used for 92% of the images, which means that the diversity of the dataset is not sufficient to evaluate its applicability to other combinations. In addition, many of the images were from the first half of the third trimester, and deep learning may be applied to MR images acquired during this period. By increasing the number of images, variation of scanner manufacturer, and magnetic field strength, deep learning predictions may gain further robustness. Second, we analyzed fetuses that had no brain abnormalities after birth, but some potential abnormalities may still have existed. Third, it is not apparent where the deep learning model was 'looking' (i.e., focusing its attention) to predict gestational age. However, all of the images used in this study were cropped around the fetal head and unified to a resolution of 150 x 150 pixels. Therefore, head size information was not directly reflected in the predictions of the model. Hence, it is likely that the model assessed brain morphology rather than size. Fourth, we only evaluated normal fetal brains and have not examined whether deep learning can evaluate abnormality in the fetal brain. Therefore, the model cannot be applied to fetuses with brain abnormalities, and also, the clinical significance of fetal abnormalities found in deep learning has not been examined. Fifth, although several parameters (such as BPD and head circumference) can assist in predicting the gestational age, it was difficult to measure anything other than BPD because of fetal movement or oblique images on MRI. A previous study has shown that BPD measured on MRI was significantly correlated with the gestational age [28]. Sixth, in this study, if the image was oblique, the image closest to the level of the thalamus and septum pellucidum was selected (with reference to the shape of the lateral ventricles). We carefully measured and tried to obtain a value close to the true BPD, but it may have been different from the BPD measured by ultrasound. Lastly, we measured BPD on MRI, not on ultrasound. This was because ultrasound examinations were not performed at the same time as the MRI examinations, making it difficult to perform clear comparisons. However, Matthew et al. [25] reported that BPD measurement on MRI produced good/excellent intraobserver agreement with higher intra-class correlation coefficients than that on ultrasound. Based on that report, we speculate that BPD prediction with MRI might be similar to that with ultrasound. It would be better if future studies could compare the results between both types of scans obtained at the same time.

In conclusion, we demonstrated that deep learning could predict gestational age from fetal brain MRI after the first trimester and could do so with high accuracy. Prediction of gestational age by deep learning may have a positive effect on prenatal care for pregnancies that are underserved during the first trimester.

Acknowledgments

We would like to thank Edanz Group (https://en-author-services.edanz.com/ac) for editing a draft of this manuscript.

References

1. Committee on Obstetric Practice, American Institute of Ultrasound in Medicine, Society for Maternal-Fetal Medicine. Committee Opinion No 700 (2017) Methods for Estimating the Due Date. Obstet Gynecol. Available via https://journals.lww.com/greenjournal/Fulltext/2017/05000/Committee_Opinion_No_700__Methods_for_Estimating.50.aspx Accessed 16 March 2020
2. Reddy UM, Abuhamad AZ, Levine D et al (2014) Fetal imaging: executive summary of a joint Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-Fetal Medicine, American Institute of Ultrasound in Medicine, American College of Obstetricians and Gynecologists, American College of Radiology, Society for Pediatric Radiology, and Society of Radiologists in Ultrasound Fetal Imaging workshop. Obstet Gynecol 123: 1070-82
3. Mongelli M, Wilcox M, Gardosi J (1996) Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. Am J Obstet Gynecol 174: 278-81
4. Bennett KA, Crane JM, O'shea P et al (2004) First trimester ultrasound screening is effective in reducing postterm labor induction rates: a randomized controlled trial. Am J Obstet Gynecol 190: 1077-81
5. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521: 436-44
6. Wang S, Summers RM (2012) Machine learning and radiology. Med Image Anal 16: 933-51
7. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing System 25: 1097-1105
8. Akkus Z, Galimzianova A, Hoogi A et al (2017) Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. J Digit Imaging 449-459
9. Gong E, Pauly JM, Wintermark M et al (2018) Deep learning enables reduced gadolinium dose

for contrast-enhanced brain MRI. J Magn Reson Imaging 48: 330-340

10. Larson DB, Chen MC, Lungren MP et al (2018) Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. Radiology 287: 313-322
11. Narayana PA, Coronado I, Sujit SJ et al (2020) Deep Learning for Predicting Enhancing Lesions in Multiple Sclerosis from Noncontrast MRI. Radiology 294: 398-404
12. Nishio M, Sugiyama O, Yakami M, et al (2018) Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. PLoS One. Available via https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0200721 Accessed 16 March 2020.
13. Cole JH, Poudel RPK, Tsagkrasoulis D et al (2017) Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. Neuroimage. 163: 115-124
14. Cole JH, Franke K (2017) Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. Trends Neurosci 40: 681-690
15. Prayer D, Kasprian G, Krampl E et al (2006) MRI of normal fetal brain development. Eur J Radiol 57: 199-216.
16. The Global Library of Women's Medicine. Assessment of Gestational Age by Ultrasound. The International Federation of Gynecology and Obstetrics Web site. Available via https://www.glowm.com/section_view/heading/Assessment%20of%20Gestational%20Age%20by%20Ultrasound/item/206. Accessed 16 March 2020.
17. Simonyan K, Zisserman A (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. Cornell University. Available via https://arxiv.org/abs/1409.1556. Accessed 16 March 2020

18. Torrey L, Shavlik J (2010) Transfer learning. Handbook of Research on Machine Learning Applications. Hershey, Pa: IGI Global, http://ftp.cs.wisc.edu/machine-learning/shavlik-group/torrey.handbook09.pdf.
19. McBride GB (2005) A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Report: HAM2005-062.
20. Nishio M, Koyama H, Ohno Y et al (2016) Emphysema Quantification Using Ultralow-Dose CT With Iterative Reconstruction and Filtered Back Projection. AJR Am J Roentgenol 206:1184-92
21. Mets OM, Willemink MJ, de Kort FP et al (2012) The effect of iterative reconstruction on computed tomography assessment of emphysema, air trapping and airway dimensions. Eur Radiol 22:2103–2109
22. Koyama H, Ohno Y, Nishio M et al (2014) Iterative reconstruction technique vs filter back projection: utility for quantitative bronchial assessment on low-dose thin-section MDCT in patients with/ without chronic obstructive pulmonary disease. Eur Radiol 24:1860–1867
23. Wu J, Awate SP, Licht DJ et al (2015) Assessment of MRI-Based Automated Fetal Cerebral Cortical Folding Measures in Prediction of Gestational Age in the Third Trimester. AJNR Am J Neuroradiol 36: 1369-74
24. Maria KM, Gerardine Q, Mary AR et al (2012) Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. Med Image Anal. 16:1550-64
25. Matthew J, Malamateniou C, Knight CL et al (2018) A comparison of ultrasound with magnetic resonance imaging in the assessment of fetal biometry and weight in the second trimester of pregnancy: An observer agreement and variability study. Ultrasound 26:229-244
26. Prayer D, Malinger G, Brugger P C et al (2017) ISUOG Practice Guidelines: performance of fetal magnetic resonance imaging. Ultrasound Obstet Gynecol 49:671-680
27. Ray GJ, Vermeulen MJ, Bharatha A et al (2016) Association Between MRI Exposure During

Pregnancy and Fetal and Childhood Outcomes. JAMA 316:952-961

28. Kyriakopoulou V, Vatansever D, Davidson A et al (2017) Normative biometry of the fetal brain using magnetic resonance imaging. Brain Struct Funct. 222:2295-2307

Table 1: Summary of main reasons for conducting fetal MRI (n = 184).

Table 2: Fetal background summary of training data, validation data, and test data.

* Mean ± standard deviation

Fig. 1: Study flowchart.

Fig. 2: VGG16 classification model modified for regression.

Fig. 3: Gestational age distribution (cases per week).

Fig. 4: Examples of brain image trimming and results of deep learning. **a** Fetal brain MR image with artifact at 14.3 weeks of gestation is predicted to be 13.0 weeks by deep learning. **b** Fetal brain MR image at 28.0 weeks of gestation is predicted to be 28.0 weeks by deep learning. **c** Fetal brain MR image at 34.3 weeks of gestation is predicted to be 34.6 weeks by deep learning. This slice was determined to be oblique.

Fig. 5: Scatterplots and regression lines showing the correlation between predictions from the model or biparietal diameter (BPD) and the reference standard. The Lin's concordance correlation coefficients of the model and BPD were 0.964 (95% confidence interval [CI]: 0.928–0.982) and 0.920 (95% CI: 0.845–0.960), respectively.

Fig. 6: Bland-Altman plots comparing gestational week predictions between the model or biparietal diameter (BPD) and the reference standard. The solid line represents the mean error, and the dashed lines represent the 95% confidence interval (CI). The deep learning model and BPD predicted the gestational age of the

fetuses after the first trimester with mean errors of −0.43 ± 1.47 (mean ± standard deviation, 95% CI: −0.99–0.13) weeks and +0.97 ± 2.37 (95% CI: 0.07–1.87) weeks, respectively. The lower limit of agreement between the model prediction and the reference standard (−3.31, 95% CI: −4.27 to −2.34) was almost the same as the lower limit of agreement between the BPD prediction and reference standard (−3.67, 95% CI: −5.23 to −2.11). The upper limit of agreement between the model prediction and reference standard (2.45, 95% CI: 1.48–3.41) was smaller than that between the BPD prediction and reference standard (5.62, 95% CI: 4.06–7.18).

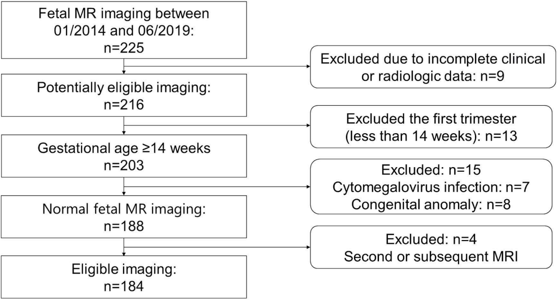
Table 1: Summary of main reasons for the fetal MRI (n=184)

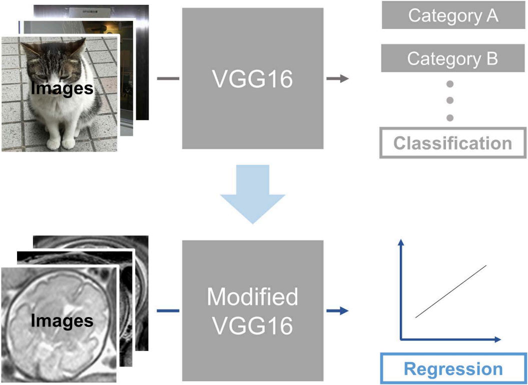
Reason for MRI	Number (%)
Umbilical cord / placental abnormalities	58/184 (32%)
Fetal abnormalities	27/184 (15%)
Pregnancy after uterine surgery	25/184 (14%)
Uterine abnormalities	20/184 (11%)
Ovarian mass	19/184 (10%)
Multiple pregnancy	9/184 (5%)
Hydramnios, Oligohydramnios	8/184 (4%)
Threatened premature delivery	8/184 (4%)
Abnormalities in pregnant women	8/184 (4%)
Other	2/184 (1%)

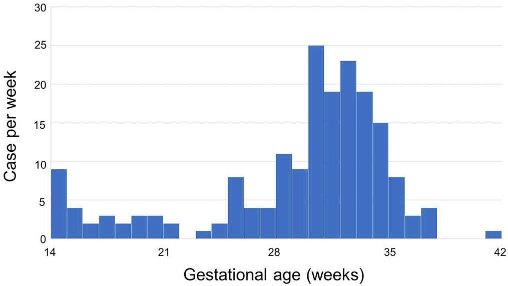
Table 2: Fetal background summary of training data, validation data, and test data.

Characteristic	Training data	Validation data	Test data
Total Number	126	29	29
Gestational Age (weeks)*	29.5 ± 6.1	29.1 ± 6.0	29.3 ± 5.8
Second trimester	31 (25%)	8 (28%)	7 (24%)
Third trimester	95 (75%)	21 (72%)	22 (76%)

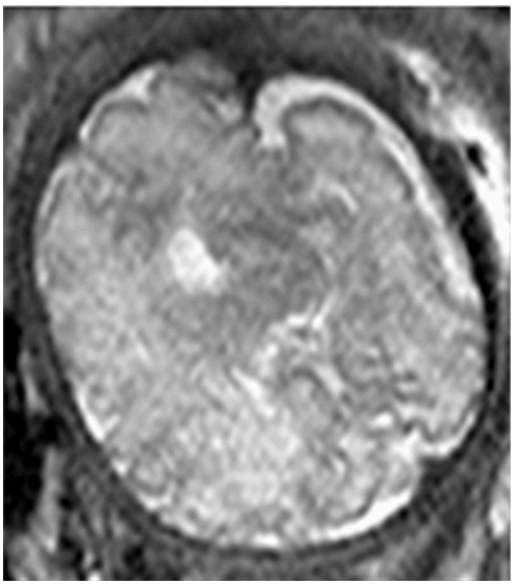
* Mean ± standard deviation

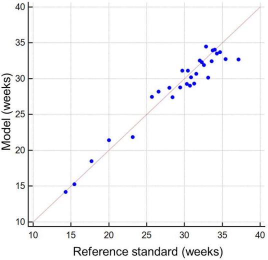

Fetal MR imaging between 01/2014 and 06/2019: n=225
Excluded due to incomplete clinical or radiologic data: n=9
Potentially eligible imaging: n=216
Excluded the first trimester (less than 14 weeks): n=13
Gestational age ≥14 weeks n=203
Excluded: n=15
Cytomegalovirus infection: n=7
Congenital anomaly: n=8
Normal fetal MR imaging: n=188
Excluded: n=4
Second or subsequent MRI
Eligible imaging: n=184


Images
VGG16
Category A
Category B
Classification
Images
Modified VGG16
Regression

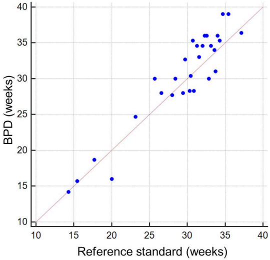


30
25
20
15
10
5
0
Case per week
14
21
28
35
42
Gestational age (weeks)

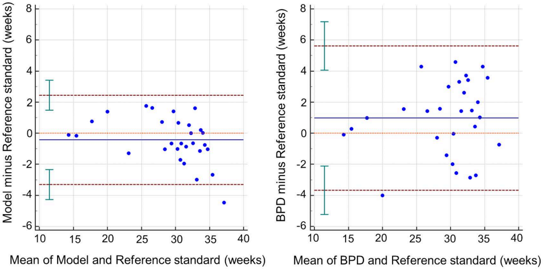




Model (weeks)
Reference standard (weeks)
10
15
20
25
30
35
40



BPD (weeks)
Reference standard (weeks)
10
15
20
25
30
35
40



Model minus Reference standard (weeks)
Mean of Model and Reference standard (weeks)
BPD minus Reference standard (weeks)
Mean of BPD and Reference standard (weeks)